# Nonverbal auditory communication – Evidence for integrated neural systems for voice signal production and perception

Frühholz, Sascha ; Schweinberger, Stefan R

Abstract: While humans have developed a sophisticated and unique system of verbal auditory communication, they also share a more common and evolutionarily important nonverbal channel of voice signaling with many other mammalian and vertebrate species. This nonverbal communication is mediated and modulated by the acoustic properties of a voice signal, and is a powerful - yet often neglected - means of sending and perceiving socially relevant information. From the viewpoint of dyadic (involving a sender and a signal receiver) voice signal communication, we discuss the integrated neural dynamics in primate nonverbal voice signal production and perception. Most previous neurobiological models of voice communication modelled these neural dynamics from the limited perspective of either voice production or perception, largely disregarding the neural and cognitive commonalities of both functions. Taking a dyadic perspective on nonverbal communication, however, it turns out that the neural systems for voice production and perception are surprisingly similar. Based on the interdependence of both production and perception functions in communication, we first propose a re-grouping of the neural mechanisms of communication into auditory, limbic, and paramotor systems, with special consideration for a subsidiary basal-ganglia-centered system. Second, we propose that the similarity in the neural systems involved in voice signal production and perception is the result of the co-evolution of nonverbal voice production and perception systems promoted by their strong interdependence in dyadic interactions.

DOI: https://doi.org/10.1016/j.pneurobio.2020.101948

Progress in Neurobiology 199 (2021) 101948

Contents lists available at ScienceDirect

Progress in Neurobiology

journal homepage: www.elsevier.com/locate/pneurobio

Review article

# Nonverbal auditory communication – Evidence for integrated neural systems for voice signal production and perception

Sascha Frühholz [a,b,c,*], Stefan R. Schweinberger [d,e,f]

[a] Department of Psychology, University of Zurich, Zurich, 8050 Switzerland
[b] Department of Psychology, University of Oslo, Oslo, 0373 Norway
[c] Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, 8057 Switzerland
[d] Department of General Psychology and Cognitive Neuroscience, Friedrich Schiller University, 07743 Jena, Germany
[e] Voice Research Unit, Friedrich Schiller University, 07743 Jena, Germany
[f] ARC Centre of Excellence in Cognition and its Disorders, School of Psychology, University of Western Australia, WA 6009, Australia

ARTICLE INFO

*Keywords:*
Communication
Voice
Auditory system
Neural network
Nonverbal

ABSTRACT

While humans have developed a sophisticated and unique system of verbal auditory communication, they also share a more common and evolutionarily important nonverbal channel of voice signaling with many other mammalian and vertebrate species. This nonverbal communication is mediated and modulated by the acoustic properties of a voice signal, and is a powerful – yet often neglected – means of sending and perceiving socially relevant information. From the viewpoint of dyadic (involving a sender and a signal receiver) voice signal communication, we discuss the integrated neural dynamics in primate nonverbal voice signal production and perception. Most previous neurobiological models of voice communication modelled these neural dynamics from the limited perspective of either voice production or perception, largely disregarding the neural and cognitive commonalities of both functions. Taking a dyadic perspective on nonverbal communication, however, it turns out that the neural systems for voice production and perception are surprisingly similar. Based on the interdependence of both production and perception functions in communication, we first propose a re-grouping of the neural mechanisms of communication into auditory, limbic, and paramotor systems, with special consideration for a subsidiary basal-ganglia-centered system. Second, we propose that the similarity in the neural systems involved in voice signal production and perception is the result of the co-evolution of nonverbal voice production and perception systems promoted by their strong interdependence in dyadic interactions.

## 1. Introduction

Communication, through which living beings signal information to and receive information from other social agents, was a catalyzer for evolution. Among the most powerful means of conveying information in vertebrate species is auditory vocal signaling and communication, up to the most evolved form of human speech and language (Hauser et al., 2002; Rauschecker, 2018; Scott, 2019). While humans would seem to communicate information mostly using verbal messages, the voice as a carrier of speech can additionally convey rich information beyond and above speech, which is highly relevant for and can directly modulate any social interaction (Argyle, 1972; Bachorowski and Owren, 2003; Dawkins and Krebs, 1978), such as emotional voice signals (Arnal et al., 2015; Panksepp, 2003; Parsons et al., 2014). This "nonverbal auditory communication" (Argyle, 1972; Rauschecker and Scott, 2009) has major communicative functions in human primates, but is critically also shared with many nonhuman species in the evolutionary linage. With the notion of "nonverbal auditory communication" we mainly refer to basic nonverbal expressions free of any speech-like content and structure, but also to acoustic vocal intonations and modulations that are superimposed on speech and speech-like material especially in humans. This latter paraverbal or paralinguistic level of nonverbal communication seems reserved for humans, but uses a similar acoustic encoding of meaning as for nonverbal expressions.

Researchers so far have focused on establishing neurocognitive and neurobiological models concerning the verbal channel for speech processing (Friederici, 2011; Hickok and Poeppel, 2007). By contrast, comprehensive models for the second channel of nonverbal auditory

* Corresponding author at: University of Zurich, Department of Psychology, Binzmuhlestrasse 14 (Box 18), 8050 Zürich, Switzerland.
*E-mail address:* sascha.fruehholz@uzh.ch (S. Frühholz).

https://doi.org/10.1016/j.pneurobio.2020.101948
Received 9 February 2020; Received in revised form 12 October 2020; Accepted 4 November 2020
Available online 12 November 2020

communication are rare and partly incomplete. Existing models also do not fully cover all aspects defining dyadic, interactive, and dynamically adapted voice signal communication (Ackermann et al., 2014; Hage and Nieder, 2016; Petkov and Jarvis, 2012). Especially, these models have mostly focused on either voice signal production or perception, without modeling these two processes in a unified theoretical approach.

In the present review, we outline a comprehensive neurobiological and neurocognitive model of nonverbal auditory communications including two major features. First, beyond a sketch of solitary processes of voice signal production in senders and voice signal perception in listeners, we understand voice communication as a dyadic interaction in a minimal sense, with mutual interactions between a sender and a listener (Stephens et al., 2010). This perspective assumes that voice communication systems did not evolve for the sender to simply express voice information, but for the effects of the perception of these vocalizations in listeners (Ehret et al., 2006). An angry vocal burst (Grandjean et al., 2005; Korb et al., 2014), for example, would be quite useless in itself, but is rather intended to cause some defensive and mindful reactions in listeners (Frühholz et al., 2015c).

Second, as we will comprehensively outline in this review, the neural mechanisms of producing voice signals share many similarities with the neural mechanisms of perceiving such signals. Voice production and perception may thus have co-evolved being mutually conditioned on each other. Efficient voice signal perception may need to consider the mechanism of how voice signals are produced (Goldman and Sripada, 2005; Niedenthal, 2007), and vice versa.

In this review, we accordingly describe the neurobiological mechanisms of voice signal production and voice signal perception. Our

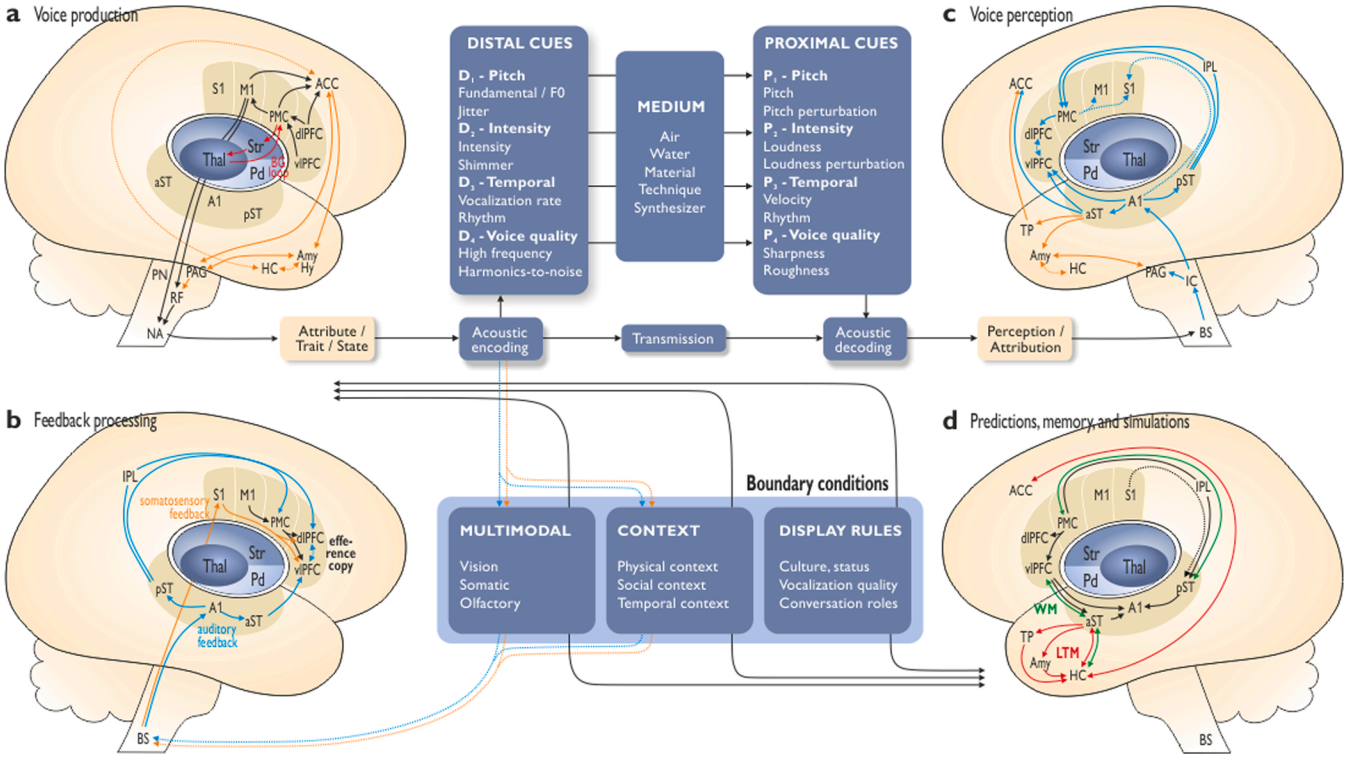


**Fig. 1.** A neural model of nonverbal auditory communication.

**(a)** Neural mechanisms during the production of a variety of voice signals. In mammals and especially in humans, voluntary control over vocal motor execution (black lines) is centered on the primary motor cortex (M1) that controls motor nuclei in the brain stem (RF, NA, motor brainstem nuclei) with direct control over NA in human primates (dashed black line). These brainstem nuclei control respiration (RF) and vocal tract musculature of the larynx (NA) and the articulators (trigeminal, facial, and hypoglossal brainstem motor nuclei). M1 is influenced by anterior regions in inferior frontal cortex (PMC, PFC), while M1 influences signal in the anterior cingulate cortex (ACC). Involuntary triggering of voice signals is driven by a second network (orange) centered on the ACC. The ACC is connected to the midbrain region PAG either directly or via the amygdala and hypothalamus. The PAG finally influences again the RF. Unlike in nonhuman primates, vocalizations in human primates potentially involve additional cortical and subcortical loops supporting vocal flexibility and learning. One important loop involves the BG system as a subsidiary system to the paramotor system, which has been centrally implicated in vocal learning in songbirds, with evidence for a potential involvement in nonhuman primates still missing.

**(b)** Voice signals produced by the sender figure as distal cues encoded in specific acoustic voice features; these features express information concerning attributes, traits, and states of the sender. These acoustic voice features are transmitted via different media and then perceived by a listener. Once transmitted, the acoustic voice features figure as proximal cues to listeners, which perceived these signals and attribute meaning to them.

**(c)** Neural mechanisms and network in listeners during the perception of voice signals. Auditory voice signals are processed along the ascending auditory pathway finally reaching the low-level (A1) and high-level auditory cortex (STS). From ST two streams emerge: an antero-ventral stream originating in anterior STS that supports the identification of different vocal information, including socio-affective meaning (Amy), sender identity (TP), and referential meaning (vlPFC), and a postero-dorsal stream originating in posterior STS for the encoding of auditory-motor information, including temporal information (PMC).

**(d)** Voice signals produced by a sender are not only perceived by listeners, but also by the sender. During the production of voice signals, senders receive two types of vocalization feedback at the auditory and at the somatosensory level. During voice production, efference copies of motor vocalization plans are stored in the frontal cortex (PMC, PFC) and incoming auditory and somatosensory information is compared against them; a difference between motor plans and feedback signals indicates vocalization errors and leads to corrections.

Abbreviations: *A1* primary auditory cortex, *ACC* anterior cingulate cortex, *Amy* amygdala, *BG* basal ganglia, *BS* brainstem, *HC* hippocampus, *Hy* hypothalamus, *LTM* long-term memory, *M1* primary motor cortex, *NA* nucleus ambiguus, *PAG* periaqueductal gray, *PN* pontine nuclei, *Pd* pallidum, *PFC* prefrontal cortex, *RF* reticular formation, *S1* primary somatosensory cortex, *Str* striatum, *ST* superior temporal cortex, *STS* superior temporal sulcus, *Thal* thalamus, *TP* temporal pole, *PMC* premotor cortex, *WM* working memory.

starting point will be current neural models of voice signal perception, such as the dual-stream model of voice signal analysis (Rauschecker, 2018; Romanski et al., 1999), models of socio-affective voice analysis (Belin et al., 2004; Frühholz et al., 2016b), and models of voice signal production (Frühholz et al., 2014a; Hage and Nieder, 2016). Given these previous models, which tend to focus only on one side of communication, we will argue towards an integrative perspective combining both production and perception aspect in nonverbal auditory communication. Our neurocognitive model of nonverbal auditory communication thus takes both the side of the sender and the listener into account (Fig. 1). While we primarily focus on human nonverbal auditory communication, we also link this perspective to primate communication in monkeys (Fig. 4) as well as to vocal signaling in songbirds (Fig. 5).

## 2. What information is encoded in nonverbal voice information?

For verbal voice communication, the nonverbal channel is a compulsory companion on which a sender volitionally (voluntarily) or spontaneously (involuntarily) sends additional nonverbal information, and on which listeners volitionally or spontaneously decode relevant information. This nonverbal channel can also be used for communication in the absence of languages, such as in nonhuman primates and other vertebrate species. To the extent that the nonverbal auditory channel is not redundant with the verbal channel, it should be of genuine relevance for any social interaction.

Although it seems difficult to give a full description and a comprehensive taxonomy of the information encoded in nonverbal voice signals, a possible five-class taxonomy might include (a) physical attributes of the sender (e.g. sex, age, height), (b) information with basic (e.g. affective state, trustworthiness) or (c) complex social information (e.g. power, competence), (d) health-related attributes (e.g. temporary or chronic diseases), and (e) non-arbitrary referential meaning to objects and environmental states (e.g. dangerous animals, delicious sweets) (Fig. 2). This taxonomy has some rationale given that this potentially relates to different dynamics in a brain network involved in their production and their perception, as discussed below.

Physical attributes of the sender comprise the first category, and they are expressed in voice signals concern features many of which are implicitly relevant for vocal social interactions, such as sex, age, height, identity, attractiveness, or mating state. These are often expressed involuntarily, but under certain conditions may be subject to signal disguise or enhancement. A suspect in a police investigation might try to change and disguise some physical voice features when vocal samples are used as evidence in an investigation (Zhang and Tan, 2008). Voice signals that indicate the sender's sex, masculinity/feminity or body size may be enhanced in the context of mating or competition (Fraccaro et al., 2011; Hughes et al., 2010; Pisanski et al., 2016). Like physical voice features, indicators of health status of the sender are also mostly expressed involuntarily, and may result from somatic and psychological disorders. A prominent example is the dysprosodic voice of patients suffering from Parkinson's disease, leading to a monotonic vocal tone (Arnold et al., 2014). While such voices may convey dynamic social information less prominently, they may provide substantial information about a sender's health status.

Basic and complex social signals are often expressed with a higher degree of intent. Basic social signals concern domains that are immediately and directly relevant for social interactions and are immediately perceived, such as vocal affect (Banse and Scherer, 1996; Frühholz et al., 2016b), trustworthiness (Belin et al., 2017; O'Connor and Barclay, 2017; Oleszkiewicz et al., 2017), sexual selection signals (O'Connor et al., 2011; Puts et al., 2012; Sulpizio et al., 2015), or linguistic/dialect affiliation (Bestelmeyer et al., 2015). Complex social information also considerably influences social interaction, but are more strongly driven by societal and cultural display rules, such as vocal signs of power and dominance (Ko et al., 2015; McAleer et al., 2014), attitude (Monetta et al., 2008; Pell, 2006), competence (Nelson et al., 2016; Oleszkiewicz et al., 2017; Sei Jin Ko et al., 2009; Tigue et al., 2012), lying and deception (Anolli and Ciceri, 1997; Rigoulot et al., 2014), sarcasm (Cheang and Pell, 2008), social status (Leongómez et al., 2017), and kinship (Levréro et al., 2015). These vocal signals can also be used in a more strategic manner in certain contexts. Politicians, for example, may modulate their voice features to sound more competent and eligible (Tigue et al., 2012).

The last category includes voice signals with non-arbitrary referential meaning to objects and environmental states. Senders may use certain voice signals and voice features to refer to the presence and location of another person or object in a certain situation. Voice signals might also index a certain environmental state, such as danger or safety in natural and societal contexts. Some nonhuman primates, for example, have voice calls to differentiate predators (Seyfarth et al., 1980) and their location (Cäsar et al., 2013). Human primates might also use a repertoire of nonverbal signals to signify the affective relevance of certain contexts (Anikin et al., 2018), such as voice signals of pre-verbal infants (Kersken et al., 2017).

Across these categories of nonverbal voice information, an important issue concerns how much each information is expressed volitionally or spontaneously by the sender, and how much of these expressions can be adapted to the context. Physical voice features and health-related

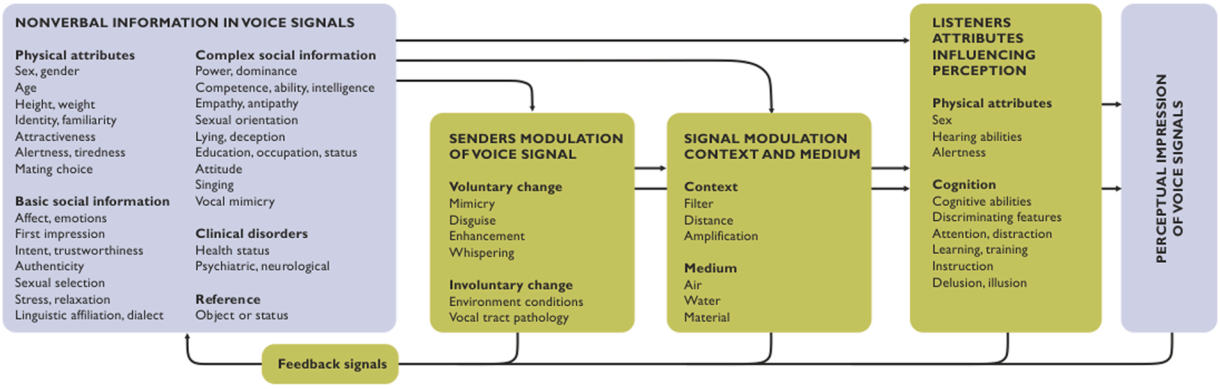


**Fig. 2.** Transmission, modulation, and perception of voice signal information.
A wealth of socially relevant information can be expressed in nonverbal voice signals (left panel) concerning physical attributes of the sender, basic and complex social information, and signs of the health and disorder status of the sender. Before voice signals are decoded and perceptually recognized by listeners (right panel), these expressed voice signals may receive some modulation and filtering by the senders, the context and medium of transmission, and the attributes of the listeners. Therefore, the signal perceived by listeners can be a degraded or modulated version of the signal expressed by the sender. The sender, in turn, may receive information about this modulation and filtering of the original vocalization by feedback signals received at several levels.

attributes, for example, concern mostly anatomical features which are largely expressed involuntarily; these can only marginally be adapted to contextual conditions. Contrarily, some complex social voice information and referential voice information is expressed volitionally to a larger degree and is often adapted to the context and to cultural display rules. A mixed mode between volitional and spontaneous expressions of voice signals concerns vocal mimicry and imitation. Infants, for example, show vocal mimicry of their caregivers (Poulson et al., 1991), songbirds imitate vocal models and tutors (Bolhuis and Gahr, 2006), and Karaoke-singers imitate the singing voice of another person (Mantell and Pfordresher, 2013), demonstrating that vocal imitation can range from spontaneous to volitional imitation. We will discuss these important topics of voluntariness, adaptation, and mimicry throughout this review.

## 3. Neurocognitive mechanisms of voice signal production

To convey voice signals, senders need to produce acoustic signals by activating the neural and anatomical apparatus of the vocal neuromotor machinery. In primates, voice signals are produced by activation of the vocal tract organ including physiological processes of respiration, laryngeal, pharyngeal, and oral motion and postural mechanisms (Fitch, 2002; Stanley, 1931). This results in a specific acoustic profile of voice signals, which we will describe in the next paragraph. To produce such acoustic profiles related to voice signals, the peripheral vocal neuromotor behavior is controlled by a neural machinery in the central nervous system, as described in the remaining paragraphs of this section.

### 3.1. The acoustic nature of nonverbal voice information

Nonverbal voice information is encoded in specific acoustic features, given that voices and voice signals are distinct auditory objects with characteristic features. During vocalizations, vocal cords vibrate with a certain frequency (fundamental frequency, f0) that mainly contributes to the perceived pitch of voice. Source signals produced by the vocal cords resonate in the oral cavity, adding formant frequencies (i.e. f1, f2, f3, etc.) to expressed voice signals. These voiced portions of vocalizations are often accompanied by unvoiced voice segments, especially when nonverbal voice signals are superimposed on unvoiced parts of verbal utterances or during whispering (Frühholz et al., 2016a). A wealth of acoustic features characterize voice signals and help to discriminate these (Frühholz et al., 2016c). These features may be grouped as pitch or spectrum related features (e.g. f0, jitter, formant distribution/timbre, spectral center of gravity), intensity-related features (e.g. intensity/energy, shimmer), temporal features (e.g. vocalization rate, rhythm), and voice quality features (e.g. harmonics-to-noise ratio (HNR), ratio of high- and low-frequency energy). Physical attributes are largely encoded in pitch and spectrum-related features of voice signals (Ghazanfar et al., 2007), such that sex is largely encoded in and decoded from voice timbre, but also from voice pitch when timbre is ambiguous (Pernet and Belin, 2012). The same seems true for the sender's height and weight (Pisanski et al., 2016; Von Kriegstein et al., 2010). Vocal attractiveness is judged on the level of pitch (Babel et al., 2014; Borkowska and Pawlowski, 2011; Fraccaro et al., 2011, 2013) and spectral smoothness of voices (Bruckert et al., 2010). The sender's age and identity is usually encoded in several combined voice features. Age is perceived from the HNR level, pitch variation and jitter (i.e. micro-fluctuations in pitch) (Harnsberger et al., 2010) as well as temporal aspects of vocalizations (Linville, 1996), while identity probably relies on the unique combination of voice features (Van Lancker et al., 1985). The latter seems also relevant for primate vocalizations, which seem to have variations between individuals (Agamaite et al., 2015) and help to differentiate between a number of conspecifics (Miller and Wang, 2006).

Unlike physical attributes, which seem to largely encode on rather static voice features, social information is largely encoded in feature variations. Signaling vocal affect includes mean level shifts and increased or decreased variation of various basic voice features (Banse and Scherer, 1996; Patel et al., 2011), such as mean and variation of the f0, intensity, and voice quality features. Trustworthy voices have a high onset f0 and finish f0 with a marked dip at mid-vocalization (Belin et al., 2017) and HNR (McAleer et al., 2014). Power and dominance is perceived from shift in the f0, the formant dispersion, and possibly the HNR (McAleer et al., 2014). Lying and deception may involve higher f0, more vocalization pauses, higher vocalization fluency (Anolli and Ciceri, 1997), marked changes in voice quality cues (Scherer et al., 1985), and potentially vocal micro tremors (Hollien et al., 1987). Sarcasm finally has a slower vocalization rate, higher voice intensity, and lower pitch (Rockwell, 2000). These few examples serve to show that the acoustic space of voice signals is large to allow signaling of rich social information, such as competence (Schroeder and Epley, 2015), power (McAleer et al., 2014), warnings and hesitation (Hellbernd and Sammler, 2016), politeness (Pell, 2007), confidence (Jiang and Pell, 2015), sincerity (Rigoulot et al., 2014), lying and deception (Anolli and Ciceri, 1997), sexual orientation (Sulpizio et al., 2015), or social status(Leongómez et al., 2017; Oveis et al., 2016). In terms of referential voice signals, some monkey species have evolved sophisticated taxonomies of alarm calls, for example, that signal the presence of certain predators (Fichtel et al., 2005; Seyfarth et al., 1980). Vervet monkeys, for example, signal the presence of eagles by low-pitched grunts, while pythons are indicated by high-pitched chutters (Seyfarth et al., 1980). Similar dedicated voice signals might indicate certain environmental states in humans that are relevant to be shared between conspecifics.

### 3.2. Neural systems for voice signal production

The neural network for production of nonverbal voice signals is composed of distributed subnetworks that subserve different functions that an organism needs to control. The major tasks for producing voice signals include executing motor behavior centered on and around the vocal tract, and leading to an acoustic appearance of a voice signal. Voice production needs to be initiated and driven by volitional and spontaneous commands to vocalize, some of which need online monitoring and fine-tuning for an accurate production. These functions are accomplished by a distributed network of brain systems.

We distinguish three major systems that contribute to the different functions of voice signal production as mentioned above, and which we describe in more detail in the following sections (Fig. 1A): (1) an "auditory (cortical) system" centered on low- (primary or core auditory cortex, A1) and high-level auditory regions in the superior temporal cortex (ST; split into anterior ST (ST) and posterior ST (pST)); (2) a distributed network covering many regions of limbic system composed of the cortical anterior cingulate cortex (ACC) and subcortical amygdala (Amy), hippocampal system (HC), and periaqueductal grey (PAG), which we refer to as the "limbic system"; and (3) a "paramotor (cortical) system" centered on primary motor cortex (M1), premotor cortex (PMC), primary somatosensory cortex (S1), and prefrontal cortex (PFC); we termed this system "paramotor system", because it is anatomically positioned largely around M1 and PMC, but it is not exclusively involved in motor functions. The "paramotor system" seems assisted by subsidiary basal-ganglia-centered system, including a neural loop between the PMC, striatum (Str), and thalamus. These systems contribute to volitional and spontaneous voice signal production as described below. They also provide differential contributions to voice signal perception and recognition as discussed further below in this review.

### 3.3. Producing different types of voice signals

Voice signal production may be volitional or spontaneous. Spontaneous voice signal production is typically elicited and largely triggered by external (e.g. approach by a predator, conspecific calls) and internal or bodily states (e.g. physical pain, mental state), and this form of voice signal production is common to many species, including human and

nonhuman primates. Unlike the spontaneous production, the volitional production of voice signal spans a broad continuum from inhibition and onset timing of vocalizations (Hage and Nieder, 2013), contextual adaptions (Roy et al., 2011; Zhao et al., 2019), interactive responding and antiphonal calling (Choi et al., 2015; Miller and Wang, 2006), up to more complex forms of volitional control such as in acting and posing (Anikin and Lima, 2018; Engelberg and Gouzoules, 2019). Most of the volitionally produced vocalizations are intended of being adapted to the current situational and conversational context. Species thus seem to differ in their ability exert basic or complex control over their vocalizations. Concerning the expression of different types of information in voice signals, during both volitional and spontaneous voice signal production (Hage and Nieder, 2016; Lauterbach et al., 2013), basic physical attributes of the sender are naturally expressed in the voice. Therefore, there are no dedicated neural systems and mechanisms for expressing physical attributes in voice signals beyond the general voice signal production machinery.

Nevertheless, some neural systems may provide supporting functions, such as when senders try to disguise or enhance physical voice attributes (Fraccaro et al., 2011, 2013). These functions critically rely both on volitional control of voice production mechanisms and on receiving auditory feedback from one's own voice. Physical attributes of senders are encoded in specific voice features, such as pitch, formant frequencies, and voice timbre, which need to be registered by the sender during own voice production. The ST, for example, shows increased activity (Parkinson et al., 2012) and left-to-right ST coupling (Parkinson et al., 2013) accompanied by increased activity on the PMC (Toyomura et al., 2007) of the paramotor system when senders recognize a pitch distortion in their voice feedback. These mechanisms might be also relevant for senders to volitionally change their physical voice features.

Volitional vocal motor behavior is centered around the paramotor system in the primate frontal cortex for motor programming commands executed by M1. M1 in humans and most likely premotor region "area 6vr" in nonhuman primates (Fig. 4A) (Petkov and Jarvis, 2012) control centrifugal efferences to brainstem nuclei that control respiration, laryngeal movement, and orofacial behavior during vocalizations. A ventral portion of the human M1 is specifically relevant for controlling laryngeal movements and is sometimes referred to as "laryngeal" motor cortex (LMC) (Simonyan and Horwitz, 2011). The human LMC or the nonhuman primate area 6vr have not been identified in other species yet (Petkov and Jarvis, 2012), but recent reports might indicate that such systems could exist also in the nonhuman primate brain (Rathelot and Strick, 2009; Rauschecker, 2018), which awaits further confirmations. Volitional production of voice signals largely concerns complex social information and signals with referential meaning. The paramotor system is involved not only in volitional preparation and execution of voice signal production, but also in the control, timing, and inhibition of more spontaneous vocalizations. In nonhuman primates, cells of the ventrolateral prefrontal cortex (vlPFC), for example, maintain prepared vocalization commands until a cue to vocalize appears. The paramotor system can also inhibit or down-regulate the urge to produce vocalizations triggered by external and internal cues (Korb et al., 2014; Lauterbach et al., 2013).

For spontaneous voice signal production that is triggered by external conditions or internal states, a pathway centered on the limbic system includes an ACC-PAG neural axis with some pathways relaying through the amygdala and hypothalamus. This pathway initiates and generates vocal patterns that are mainly driven by intrinsic motivational and emotional factors (Hage and Nieder, 2016; Jürgens, 2002). While activity in the PAG is more tightly coupled to vocalizations with latencies less than one second, activity in the ACC is less tightly linked to vocalizations in a temporal sense (Jürgens and Ploog, 1970).

The volitional pathway at the level of the paramotor system can feed information into the limbic pathway, but can also inhibit the limbic pathway at the level of PAG projections to the brainstem (Lauterbach et al., 2013). Dysfunctional inhibition may lead to involuntary bursts of affective vocalizations, such as in pathological crying and laughter (Lauterbach et al., 2013). This spontaneous production of voice signals mainly concerns the expression of simple social information, such as vocal affect, stress, trustworthiness, and authenticity. The expression of spontaneous laughter in humans comprise areas of the amygdala, hypothalamus, and PAG (Wattendorf et al., 2013), and a negative association between the ventral PFC and the limbic system (i.e. ACC and amygdala) during vocal stress (Laukka et al., 2011) (see Fig. 6A for recent studies on functional brain connectivity during voice signal production). Voluntary laughter additionally activates Heschl's gyrus of the auditory cortical system (Wattendorf et al., 2013), highlighting the relevance of auditory feedback processing. A combined involvement of the paramotor, auditory, and limbic system was also observed for the voluntary expression of vocal anger (Frühholz et al., 2015c; Klaas et al., 2015) as well as for other expressions of affect (Pichon and Kell, 2013).

While simple social information is usually expressed spontaneously, these vocalizations can also be expressed voluntarily and strategically in primates in certain contexts. Humans may have learned via operant conditioning that certain expressions have desirable effects in certain contexts, or with certain listeners (Scherer, 1986). Nonhuman primates can also make strategic use of their vocalizations (Silk et al., 2016) and adapt them to the context (Clarke et al., 2015) by modulations of certain voice features (Hage and Nieder, 2016; Petkov and Jarvis, 2012). Such strategic expression of voice signals likely involves the paramotor system (Hage and Nieder, 2016) next to the limbic system. Evidence for the relevance of the paramotor system in strategic expressions in nonhuman primates comes from animal research including old-world (macaques) (Hage and Nieder, 2016; Petkov and Jarvis, 2012) and new world monkeys (marmosets) (Miller et al., 2015, 2010; Roy et al., 2016). Strategic expressions may also involve the BG system involved in vocal learning and the representation of vocal habits, which is an important subsidiary function of the paramotor system.

### 3.4. Learning to produce voice signals

Next to volitional and spontaneous limbic vocal motor pathways, human voice signal production probably involves two additional neural systems. The neural system discussed in the previous section might be sufficient for the volitional and spontaneous production of relatively simple and short voice signals. However, some human expressions of voice signals can be of a more complex vocal and temporal nature, and their expression may be influenced and learned by social groups and contexts. This issue seems especially relevant for voice signals that are simultaneously expressed with and superimposed on verbal utterances. Speech prosody is the carrier of such nonverbal voice information, and speech utterances impact on the complexity of nonverbal voice information in speech prosody. This seems to mainly apply to more complex vocal emotions, such as shame, guilt, or pride, that usually have are of more extended vocal durations and vocal dynamics (Alba-Ferrara et al., 2011).

More complex nonverbal voice signals both follow more "innate" and anatomical patterns that drive voice production and are additionally governed by mechanisms of vocal learning. Vocal learning includes the ability to control and adjust the production of vocalizations supported by more sophisticated vocal programming. For example, affective prosody (Banse and Scherer, 1996) – the suprasegmental modulation of the speech melody to express the sender's affective state – involves a temporal pattern of vocal tone modulations that can support the verbal message, but can also directly impede the dynamics of social interaction. Although affect in speech prosody can be driven by involuntary voice patterns, affective prosodic intonation undergoes vocal learning and programming in ontogenetic development (Baltaxe and Simmons, 1985). This vocal learning and programming may be accomplished by the basal ganglia (BG) system. The BG system plays an important part in motor and habit learning (Ashby et al., 2010), and is also relevant for vocal learning in humans (Jarvis, 2007; Simmonds et al., 2014).

This contribution and relevance of the additional BG system to vocalizations and vocal learning, however, is very different across species. This notion points to an important distinction (or continuum (Petkov and Jarvis, 2012)) about the species classified as being rather "vocal learners" or rather "vocal nonlearners" (Petkov and Jarvis, 2012). This distinction or assumed continuum seems an interesting concept, but the exact neural mechanism underlying this distinction needs further empirical investigations. On the phenomenological level, vocal nonlearners concern species (including nonhuman primates) that show some limited and mostly inborn vocal repertoire. Nonlearners' voices are largely used in a prototypical expression mode with only minor vocal variations, and are triggered by situational and contextual factors, although some voice features may be adapted to the context (Egnor et al., 2006; Hage and Nieder, 2016; Petkov and Jarvis, 2012). Unlike vocal nonlearners, vocal learners (including human primates) show developmental and experience-dependent changes in their vocal expressions and vocal modulations based on model and tutor imitation, auditory learning, and vocal practice. They usually show a more flexible nonverbal communication repertoire and fewer innate auditory voice templates (Kroodsma and Pickert, 1984).

These vocal learning processes, which are often preceded by a perceptual auditory learning phase without any active production of vocalizations in some species (Bolhuis et al., 2010; Doupe and Kuhl, 1999), require additional neurocognitive processes such as auditory memory, vocal error monitoring and online fine-tuning, which seem supported by the BG (Bolhuis and Gahr, 2006). Activity in the BG is accompanied by activity in other neural systems, such as the cortical association areas of the auditory system for vocal memory (Bolhuis and Gahr, 2006; Frühholz et al., 2015c) and the paramotor system for monitoring of ongoing vocalizations (Frühholz et al., 2015c; Klaas et al., 2015). The BG system might support the learning of vocal motor programs and routines, induce vocal tone modulations for signaling (Caekebeke et al., 1991; Van Lancker Sidtis et al., 2006), and online evaluations and error detection for own vocalizations together with the paramotor system (Bolhuis et al., 2010).

Human patients with BG lesion have difficulties expressing social information in the vocal tone based on the inefficiency of using vocal tone modulations (Arnold et al., 2014; Möbes et al., 2008; Van Lancker Sidtis et al., 2006) to produce acoustically rich vocalizations. The BG system and circuit might be especially relevant for vocal mimicry and imitation (Bolhuis and Moorman, 2015; Frühholz et al., 2015c) as important social forms of vocal learning. These learning and imitation mechanisms critically depend on registering and evaluating the sender's own vocal performance based on auditory and somatosensory feedback.

An important notion about the role of the BG on vocal learning and its different relevance across species especially concerns the current evidence in nonhuman primates. Nonhuman primates are often classified as "vocal nonlearners" (Petkov and Jarvis, 2012), although some form of vocal modulations and adaptions are evident in these species, such as controlling the timing of vocalizations to cues (Hage and Nieder, 2013) and to avoid interference with external noise (Roy et al., 2011) or interfering sounds (Zhao et al., 2019), or the development of population-specific dialects (De La Torre and Snowdon, 2009; Zürcher et al., 2019). This points to some basic and rather long-term, but potentially feedback-related vocal plasticity in primate species, which might point to some involvement of the BG system in these species next to a central role of the auditory cortex (Eliades and Tsunada, 2018). Although as yet no strong evidence for an involvement of the BG circuit in nonhuman primates exists (Rauschecker, 2018), note that this could be largely due missing investigations in this topic. Thus, the current picture on the involvement of the BG system in primate vocalizations might be described as the absence of evidence rather than a strong evidence of absence about the BG relevance in these species. Future studies might look deeper into this pending scientific question.

### 3.5. Registering own voice signals by senders

Next to these major neural systems and circuits for voice production, own voice signals are registered by the sender by means of two important feedback circuits. The sender can listen to own vocalizations by means of feedback that is mediated by the auditory system (Brainard and Doupe, 2000a), and can also sense own motor behavior by means of somatosensory feedback registered in the S1 as part of the paramotor system (Hickok et al., 2011; Houde and Chang, 2015; Tremblay et al., 2003; von Holst and Mittelstaedt, 1950). Although all species that produce vocalizations can probably register the feedback from own voice signals, the relevance of this feedback for ongoing and future vocalizations might be less universal. Vocal nonlearners, such as nonhuman primates, might register feedback from their own vocalizations (Eliades and Wang, 2008), but this feedback seems relatively irrelevant for vocal plasticity beyond simple voice feature adaptations (Egnor and Hauser, 2004).

In vocal learners, vocal motor commands executed by M1 are temporarily stored as an efference copy in the anterior paramotor system (Houde and Chang, 2015). Subregions located in the anterior paramotor system (BA45, BA6) also shows preparatory activity before vocal onset, and its activity is associated with certain produced voice features (Hage and Nieder, 2013). Online (Frühholz et al., 2015c; Pichon and Kell, 2013) and offline listening (Kaplan et al., 2008) to own voice signals in senders usually activates the anterior paramotor system. Incoming own-voice feedback and somatosensory signals are compared against this efference copy and assessed for differences between the vocal plan and its realizations, and accordingly used for vocal adaptations (Houde and Chang, 2015). Differences between produced voice signals and the efference copy might lead to vocal adaptations of ongoing or future vocalizations. These vocal adaptations especially concern the volitional expression of voice signals, but they might be also relevant for spontaneous vocal expressions (Eliades and Wang, 2008). Senders might realize, for example, that a spontaneous affective burst might have been interrupted, inappropriately expressed, or too weakly expressed in a certain context, and this might lead to more adaptive future affective bursts.

A common observation is that produced vocalizations that meet the requirements of the predicted vocal intention stored as efference copy lead to a decreased signal in the auditory system indicating a successful vocal expression (Creutzfeldt et al., 1989; Toyomura et al., 2007). While this observation has been mainly made with verbal signaling (Hickok, 2012) and signaling physical attributes of senders, such as vocal pitch (Behroozmand et al., 2009; Toyomura et al., 2007), studies on signaling basic and complex social information observed rather increased auditory system activity for successful voice signaling (Frühholz et al., 2015c; Wattendorf et al., 2013). It seems like the increasing social relevance of voice signaling attract auditory processing resources even for successful vocalizations.

This increased auditory activity might play a role either in auditory memory for own and other voice signals (Bolhuis and Gahr, 2006), for compensatory plasticity effects induced by long-lasting vocal production deficits (Cheung et al., 2005), the imagination of own voice signals (Alderson-Day et al., 2015), and could be also based on an additional involvement of the limbic system for own production and perception (Frühholz et al., 2015c; Pichon and Kell, 2013; Wattendorf et al., 2013). This involvement of the limbic system seems to drive the increased activity for own voice social signal perception.

## 4. Neurocognitive mechanisms of voice perception

Voice signals are transmitted to listeners by different means and media (Figs. 1, 2, and 3). When reaching listeners, acoustic features of voice signals figure as proximal cues that are perceived, decoded, and eventually recognized and categorized. As noted above, the neural machinery for voice perception seems to involve similar neural systems

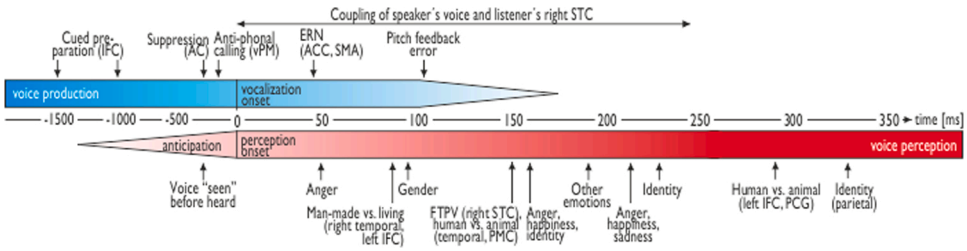


**Fig. 3.** Time course of voice signal production and perception.
Behavioral and electrophysiological data point to the temporal dynamics of producing voice signals in senders (blue stream) and of perceiving voice signals in listeners (red stream). Time point zero on the time axis denotes the onset of the vocal signal production, which is preceded by preparatory neural activity especially in the IFC (Hage and Nieder, 2013) around 1500–1000 ms prior to voice onset (i.e. vocalizations in head-restrained monkeys) as well as ~200 ms in PMC(Miller et al., 2015) (i.e. in freely moving monkeys). After the onset of voice signal production, activity in the motor system (error-related negativity, ERN(Masaki et al., 2001)) and the auditory system signify vocalization success or errors (e.g. pitch feedback errors(Behroozmand et al., 2009; Korzyukov et al., 2012)). After vocalization onset the neural processes in the listener's brain (i.e. in right ST) couple with voice features of the speaker's voice (Bourguignon et al., 2013). On the side of the listener, early neural responses start a little before voice onset, such that listeners anticipate the appearance of voices based on initial face movements leading to the phenomenon that voices are "seen" before actually heard(Joosten et al., 2015; Sodoyer et al., 2009). After voice onset, certain voice signal information can be decoded at different latencies. Major distinctions of voice signal information from other sounds appear at early (man-made vs. living sounds ~90 ms)(Murray et al., 2006), mid (voice vs. other sounds, fronto-temporal positivity to voices, FTPV(Capilla et al., 2013), ~150 ms; human voice vs. animal sounds(De Lucia et al., 2010), ~150 ms), and late response latencies (human voice vs. animal sound, ~290 ms)(Murray et al., 2006). More specific voice information is also decoded at different latencies, such that emotions affect neural processing(Brosch et al., 2009; Jessen and Kotz, 2011a; Paulmann and Kotz, 2008; Pell and Kotz, 2011; Schirmer et al., 2013; Schirmer and Escoffier, 2010; Wambacq et al., 2004) at early (anger ~50 ms), mid (anger, happiness ~160 ms, other emotions ~190 m), and late latencies (sadness ~210 ms, anger, happiness at ~220 ms). Physical features of the speaker's voice (Schweinberger et al., 2014, 2008; Zäske et al., 2009) are decoded also in early (gender ~90 ms), mid (identity ~160 ms), and late latencies (identity ~230 ms and ~330 ms); the latter showed scalp surface event-related potential (ERP) effects over temporal cortex that might be source-localized to specific brain regions in future studies.

as for voice signal production, comprising again the three broad systems described above. However, these systems have differential functions to support the four major tasks involved in voice perception: (a) bottom-up acoustic analysis, temporal decoding, and classification, (b) socio-affective analysis, (c) top-down predictions about the potential voice signal category and identity based on contextual factors and cued memory associations, and (d) behavioral preparations to adaptively respond to perceived voice signals.

### 4.1. *Perceiving and decoding voice signals*

In listeners, the auditory processing of voice signals of senders includes similar mechanisms in the auditory system as for general sound processing. This processing is also similar to the voice feedback processing of own-voices in senders as described above, but with a differential connectivity to other supporting systems.

After some detailed acoustic analysis of voice signals in the primary and secondary (core and belt) auditory cortex (Griffiths and Warren, 2002; Leaver and Rauschecker, 2016, 2010) and acoustic feature

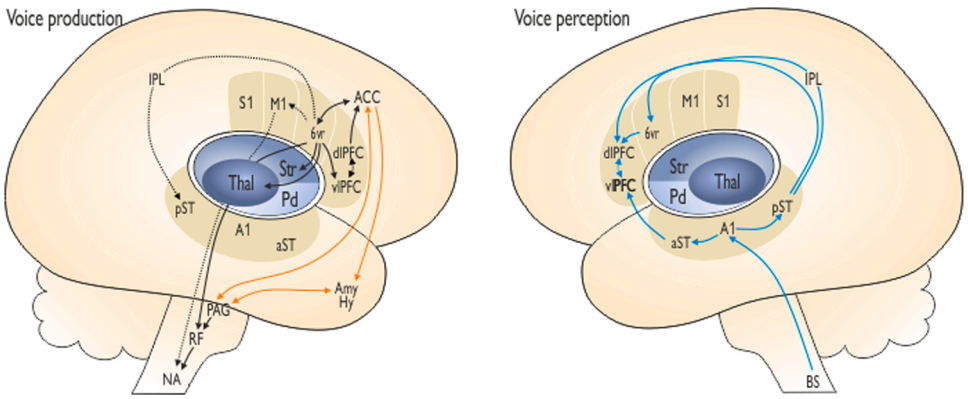


**Fig. 4.** Neural mechanisms of voice signal production and perception in nonhuman primates.
Nonhuman primates are phylogenetically the closest relatives to humans and share nonverbal voice signaling with humans but, unlike humans, are largely vocal nonlearners. Thus, nonhuman primates usually have a limited repertoire of nonverbal voice signals that are barely plastic and only have very limited flexibility. They can control the timing of vocalizations (Hage and Nieder, 2013), but have only rudimentary capacities for acoustic modulations of voice signals (Hage et al., 2013). Nevertheless, nonhuman primates show nonverbal voice signals that resemble in their acoustic nature some voice signals in humans (Scheumann et al., 2014) Neural mechanisms of voice signal production in nonhuman primates (left panel) include two major neural systems(Hage and Nieder, 2016). The system for spontaneous voice production is similar to the one in humans and comprises regions of the limbic system including the ACC-PAG axis. This system is especially relevant in nonhuman primate communication, since most of their vocalizations are driven by socio-affective triggers. The second system for more voluntary voice signal production is partly similar to, but less complex than, the paramotor system in humans. This system comprises areas in the frontal cortex including the PFC (ventral and dorsal), and a premotor area termed "area 6vr" representing the rostral part of PMC. Area 6vr has projections to subcortical structures of the BG system, and to the brainstem RF. Unlike humans, no direct projections from the paramotor system to the NA are known in nonhuman primates (see however a recent report pointing to some direct projections (Rathelot and Strick, 2009)). Furthermore, no strong evidence for an involvement of the BG circuit in nonhuman primates exists so far (Rauschecker, 2018). Voice perception (right panel) in nonhuman primates involves the auditory system and ventral and dorsal projections to the paramotor system.

integration in higher-level ST (Belin et al., 2004; Jasmin et al., 2019; Kumar et al., 2007; Leaver and Rauschecker, 2010), recent models suggest two streams to feed-forward information from the auditory belt to the paramotor system of the IFC. While a ventral stream from the anterior ST to the ventrolateral PFC is involved in sound category and identity recognition and indexical referential meaning decoding, a dorsal stream from the posterior ST to PMC has been suggested to analyze auditory space and motion (Averbeck and Romanski, 2004; Romanski et al., 1999), to provide a sensorimotor based auditory-to-motor mapping (Rauschecker, 2012, 2011; Rauschecker and Tian, 2000), and a hierarchical coding of temporal sound sequences (Bornkessel-Schlesewsky et al., 2015). These pathways have been identified for both nonverbal voice perception and speech recognition (Friederici, 2012; Hickok and Poeppel, 2007), emphasizing their general importance for vocal communication (Rauschecker and Scott, 2009).

These two streams for general sound analysis appear to be complemented by a dedicated network for the processing of voice signals. In particular, specialized auditory cortical circuits in the ST are more sensitive to voices compared to other sounds signals. These bilateral cortical areas have been termed "temporal voice areas" (TVA), and were identified in human (Belin et al., 2000; Pernet et al., 2015) and nonhuman primates (Belin et al., 2018; Perrodin et al., 2011; Petkov et al., 2008; Sadagopan et al., 2015) and in dogs (Andics et al., 2014). In humans, these areas develop around 7 months of postnatal age (Blasi et al., 2011; Grossmann et al., 2010), in parallel with adult-like processing of naturalistic sounds in the auditory system by 3−9-month old infants (Wild et al., 2017). The TVA not only responds stronger to voice compared to other sounds but some of its subregions show enhanced activity to various types of voice signals, such as voice identity (Latinus et al., 2013), gender (Weston et al., 2015), body size (Von Kriegstein et al., 2010), affect (Ethofer et al., 2012; Frühholz and Grandjean, 2013a), attractiveness (Bestelmeyer et al., 2012), dialect and accent (Bestelmeyer et al., 2015). In deaf individuals, the TVA may even support visual face processing, demonstrating functional plasticity (Benetti et al., 2017).

It seems, therefore, that the auditory cortical system is suitable for the decoding of a wealth of voice signals. Beyond this sensitivity to a multitude of voice signals, however, no clear topological structure of separate subregions uniquely sensitive to specific voice signal information has emerged. Beyond the TVA, there are additional regions that decode specific voice information, such as voice identity in the temporal pole (TP) (Perrodin et al., 2015). Although the TP is part of the temporal

neocortex, it has been proposed to be associated with the limbic system by virtue of its anatomical connectivity (Amaral and Price, 1984; Olson et al., 2007).

Voice identity recognition might involve socio-affective evaluations since identity is probably stored based on the socio-affective relevance of other individuals, as found in facial communication (Ellis et al., 1997; Schweinberger and Burton, 2003). More generic vocal socio-affective information and social associations are also neurally decoded in the amygdala, TP, and also the PAG (Dricu et al., 2017; Dricu and Frühholz, 2016; Frühholz et al., 2016b, 2014b; Frühholz and Staib, 2017; Pannese et al., 2016). The latter, for example, is active in adults listening to infant distress vocalizations (Parsons et al., 2012). Mentalizing and complex social judgments of voice signals may involve the ACC as part of a broader medial frontal area, which is often found on social mentalizing tasks (Amodio and Frith, 2006), and which receives voice information from ST and TP via uncinate fibers (Muñoz et al., 2009; Petrides and Pandya, 1988).

A critical and widely debated question is if different subregions of the paramotor system are also involved in voice signal perception. This notion seems odd at first since there is no obvious reason that a motor system should be involved in sensory perception. But the proposal of "mirror" neurons (Rizzolatti and Craighero, 2004), which are active when individuals passively perceive actions of others, has inspired some researchers to propose that listeners engage their motor and premotor system to facilitate the recognition of socially relevant expressions (Goldman and Sripada, 2005; Niedenthal, 2007). Receivers activate their (pre-)motor system to decode facial expressions of others, and they might also do so for decoding information from sound (Gazzola et al., 2006) and specifically from voice signals (Warren et al., 2006). Knowing and simulating how something is produced might critically facilitate the perception of such signals. Listeners might also recognize and mirror somatosensory processes in senders (Keysers et al., 2010; Kragel and LaBar, 2016), which again might facilitate an understanding of voice signals.

Recent studies demonstrate motor cortex involvement in perceiving verbal voice signals (Cheung et al., 2016; D'Ausilio et al., 2009; Schomers and Pulvermüller, 2016), and the motor and premotor system might support voice perception for complex voice signal patterns (Williams and Nottebohm, 1985), such as affective vocalizations (Warren et al., 2006) and affective prosody (Frühholz et al., 2016b). Next to M1, primary somatosensory cortex (S1) also seems to support voice signal perception. S1 is able to distinguish between affective categories (Kragel and LaBar, 2016), which might support the decoding of social voice information. This might be based on auditory frequency representations in human S1 (Pérez-Bellido et al., 2018) and probably also on superior colliculi (SC) connections to the auditory system (Smiley and Falchier, 2009).

The neural machinery for voice perception not only includes the auditory, limbic, and paramotor system, but critically also the BG system as an important subsidiary system to the paramotor system. While the BG system might be assumed to be primarily supporting voice production given its strong association with the paramotor system, recent evidence points to its critical involvement also in voice perception (Frühholz et al., 2016b). While the prominent role of the BG system in voice production is associated with different functions of voice motor learning and plasticity (Kojima et al., 2013; Ziegler and Ackermann, 2017) and with voice modulations (Walsh and Smith, 2012), its role in voice perception seems tightly linked to the temporal dimension of voice signals. The BG system supports temporal decoding, binding, the anticipation of sound (Geiser et al., 2012; Leaver et al., 2009) and especially of vocal events and patterns (Kotz et al., 2009; Kotz and Schwartze, 2010). This might be accomplished by its connection to the auditory system (Yeterian and Pandya, 1998) (see Fig. 6B for recent studies on functional brain connectivity during voice signal perception). This might involve decoding of temporal variations and patterns, such as in affective prosody to facilitate the decoding of emotional meaning (Frühholz et al., 2018; Grahn and Brett, 2007; Hass and Herrmann, 2012). Lesion in the BG system can impair processing of socio-affective voice information, especially at late cognitive stages of processing (Paulmann et al., 2011).

### 4.2. Predictions, simulations, and memory

The neural system for voice perception not only passively registers and analyses signals, but it actively configures and prepares the neural network for voice perception (Fig. 1D). The situational (i.e. in specific situations, certain vocal signals are more likely than others) and temporal context of a voice signal (i.e. based on preceding vocalizations and interactive behavior, certain vocal signals are more likely) that is common to the sender and the listener as well as the listeners' previous experiences may provide cues to the proper production and understanding of voice signals. These contextual cues allow predictions about meaning and social relevance of vocalizations, respectively. These predictions are most likely generated in the anterior PFC of the paramotor system (Rauschecker and Scott, 2009) and probably also in the S1 (Smiley and Falchier, 2009), and are mapped forward to auditory cortex regions that are involved in voice signal and socio-affective analysis at multiple processing levels. These predictions are assessed against incoming information; if disconfirmed, surprise and additional signal analysis or re-adjustment of predictions are initiated (Friston and Frith, 2015; Mechelli et al., 2003).

Perceptual adaptation in voice perception (Schweinberger et al., 2008) seems to be a common mechanism that exemplifies the role of predictions. Essentially, adaptation to different cues in voices (such as speaker or speech identity) elicits specific decreases in neural responses to these cues in auditory cortex areas (Belin and Zatorre, 2003). Voice adaptation during voice perception has recently been demonstrated for a number of social cues about a speaker's gender (Schweinberger et al., 2008; Zäske et al., 2009), age (Zäske and Schweinberger, 2011), identity (Latinus and Belin, 2011; Zäske et al., 2010), or emotional state (Bestelmeyer et al., 2014). Moreover, adaptation can modify voice perception even cross-modally, such that silent facial videos with strong social cues can systematically bias the subsequent perception of voice signals (Pye and Bestelmeyer, 2015; Skuk and Schweinberger, 2013). This points to the important notion of a rather multisensory nature of nonverbal communication, which we will discuss below.

Predictions during the perception of voice signals cannot only be made intra-modally but also based on information from other sensory modalities. For example, silent lip-reading affects activity in auditory cortex (Calvert et al., 1997), and visual facial information can help to predict, categorize, or even disambiguate voice information. The brain specifically exploits previously encoded audiovisual correlations to optimize communication by simulation of talking faces (Von Kriegstein et al., 2008).

While internal prediction based on multimodal, contextual, and temporal expectations usually facilitate voice perception, strong and dysfunctional predictions might also lead to misperceptions and delusions of voice signals. In extreme conditions, this can also lead to the illusion of hearing voices without any sensory stimulation, such as found in some psychotic disorders leading to voice signal illusions (Hugdahl, 2009). Some patients hear voices in the absence of external voice signals, probably by overemphasizing internal acoustic representations (Ford et al., 2009) that might be caused by an impairment of synchronizing pre-vocalization frontal activity that usually suppresses auditory evoked responses while listening to own vocalizations (Ford et al., 2007)

Voice signal decoding cannot only be actively influenced by predictions and expectations based on external cues but also based on internal cues related to memory processes and previous experiences stored in short-term working memory (WM) (Kumar et al., 2016; Pasternak and Greenlee, 2005; Scott et al., 2014) or long-term memory (LTM)(Suga et al., 2004). WM processes are centered on the paramotor system's interaction with the auditory system (Constantinidis and Procyk, 2004;

Kumar et al., 2016; Lemus et al., 2009; Munoz-Lopez et al., 2010; Scott et al., 2014), and may influence the current perception of voice signals based on representations of, or computations on, preceding voice signal information. The meaning of current socio-affective voice signals, for example, is partly influenced by the temporal sequence of preceding socio-affective signals (Mitchell, 2007).

Auditory LTM processes are centered around the connectivity of the auditory system (Suga et al., 2004) to parts of the limbic system, especially to the amygdala-hippocampus axis (Frühholz et al., 2014b) and some other regions of the medial limbic system (Munoz-Lopez et al., 2010). These LTM processes may contribute episodic voice memory information and general voice knowledge to the current perception of voice signals, but also in case voice signals do not carry enough information for a proper recognition, such as in whispering (Frühholz et al., 2016a). LTM information can also contribute to the imagination and internal simulation of voices as discussed before. Listeners not only simulate voice perception triggered by external cues but also as a process to imagine previous encounters and experiences based on simulations and contextual replay referred to as "situated conceptualizations" (Barsalou, 2009).

## 5. The interplay between production, transmission, and perception

In the previous sections, we discussed the neural, cognitive, and contextual dynamics of producing, transmitting, and perceiving voice signal information. We also outlined the neural machinery that supports a sender's voice production and the neural machinery that supports a listener's voice perception. Intriguingly, the neural systems supporting voice production and perception, at least in humans, show high similarity across the three major systems (auditory, limbic, and paramotor/BG) described above. There are potentially two major explanations for this neural overlap referred to, first, as the multifunctional perspective of the brain and, second, as the integrated functioning of semi-specialized production-focused and perception-focused systems both in senders and listeners, respectively.

### 5.1. The multi-functional perspective

The first potential explanation is referred to as "multi-functional perspective" and revolves around the long-lasting debate of functional specialization of higher-level, but also of low-level brain areas. This perspective includes the central hypothesis that a highly specialized region would serve only one single neurocognitive function. Rather than a full specialization of certain brain regions as the extreme version of this hypothesis, we might think of brain areas that are dedicated to a domain of functions instead of one singular function.

For example, the domain of functions ascribed to the auditory system might not only comprise the functional processes of sensory analysis of acoustic information as part of voice signal perception but also the acoustic preparation of vocalizations as part of voice production, such as the representation of vocal templates to prepare subsequent vocalizations (Arnold et al., 2014; Pichon and Kell, 2013). Similar to vocal preparation in humans, vocal templates represented in auditory association cortex support song production in songbirds (Bolhuis and Gahr, 2006; Bolhuis and Moorman, 2015; Hahnloser and Kotowicz, 2010).

Thus, the auditory system is also involved in voice production and specifically in vocal motor preparation. In turn, there is also evidence for an involvement of the paramotor system, and especially M1 in voice perception. M1 has been found to respond to perceived voice signals and shows an auditory rather than a motor gradient mapping to the perceived voice signals (Cheung et al., 2016), demonstrating its sensitivity to perceived acoustic voice information beyond its central role in vocal motor behavior.

Next to the auditory and paramotor systems, the limbic system also shows strong overlap in voice signal production and perception. The amygdala, the PAG, and the ACC are involved both in socio-affective voice signal production (Frühholz et al., 2015c; Wattendorf et al., 2013) and perception (Frühholz et al., 2016b; Frühholz and Grandjean, 2013b; Parsons et al., 2012). Limbic dysfunctions can impair the production (Lauterbach et al., 2013) and perception (Frühholz et al., 2015b) of socio-affective voice information, and this can be an indicator of the health and clinical status portrayed and perceived in voice signals.

Additional to the limbic system, the BG system is also involved in voice signal production and perception (Frühholz et al., 2016b; Kotz et al., 2009; Van Lancker Sidtis et al., 2006), but is differentially connected with the primary systems responsible for production or perception, respectively. While the BG system is functionally interconnected with the paramotor system during voice signal production (Bolhuis et al., 2010; Klaas et al., 2015; Petkov and Jarvis, 2012; Pichon and Kell, 2013; Ziegler and Ackermann, 2017), it is functionally connected with the auditory system during voice signal perception (Abrams et al., 2016; Frühholz et al., 2016b). This observation exemplifies the important notion that while many brain systems might have a multi-functional role during voice signal production and perception, the relative contribution to each process may be different. Subregions of higher-level auditory cortex as part of the auditory system represent the central neural node for voice signal perception, while the paramotor system and the limbic system are the central neural nodes for voice signal production. Other systems may differentially contribute to each domain of voice production and perception.

### 5.2. The integrated functioning perspective

The "integrated functioning perspective" involves the hypothesis that accurate functioning within one domain (e.g., voice perception) largely relies on the other domain (e.g., voice production). This implies that each neural system provides selected rather than multi-functional processes for voice production and perception, but that we simulate the other process (e.g. motor functions are simulated during voice perception) while executing the primary one. This may be referred to as a "simulationist" model. This model has become prominent for perceiving expressions (Goldman and Sripada, 2005; Niedenthal, 2007), with the notion that one domain does not properly function without crucial contribution by the other domain.

This view is supported by observations that patients with deficits in the production of socio-affective information also tend to be impaired in the recognition of such information (Niedenthal, 2007). Parkinson patients, for example, have a primary deficit in producing socio-affective voice information (Arnold et al., 2014; Caekebeke et al., 1991; Sidtis and Van Lancker Sidtis, 2003), but also show impairments in socio-affective voice perception (Péron et al., 2012). Until now, this hypothesis has been only formulated for the subsidiary use of the paramotor functions for the primary process of voice signal perception. Voice signal perception is accordingly facilitated if a listener simulates (Goldman and Sripada, 2005; Niedenthal, 2007) how this signal has been produced (Lindblom, 1996). This seems especially relevant when listeners aim at imitating a new vocal signal. This motoric and somatosensory replay in listeners could promote an embodiment of the sender's subjective states that most likely drove their vocal signal expressions (Niedenthal, 2007), which may facilitate the understanding of voice signals.

So far, the integrated functioning perspective mainly discussed the involvement of paramotor process in voice signal perception. Explicit hypotheses of an involvement of perception mechanisms, especially represented by the auditory system, on voice production are comparatively scarce. Recent empirical findings could support the notion that perceptual mechanisms may contribute to voice signal production. For example, the production of socio-affective voice signals seems overall only partially correlated with the physiological and neural affective state of the sender (Bachorowski and Owren, 2003). To the extent that voice signal production is not exclusively shaped by the sender's state,

the anticipation of a listener's perception of these voice signals and the effects in the listener may contribute to voice signal production in senders (Bachorowski and Owren, 2003).

Voice signals produced by some animal species are assumed to be produced to manipulate the state of the listener, rather than primarily expressing information by the sender (Dawkins and Krebs, 1978). The distress call of infant guinea pigs during social separation, for example, is highly efficient in attracting the attention of caregiving guinea pigs (Panksepp, 2003). This does not necessarily imply that a sender explicitly anticipates the perceptual effects in listeners, since successfully influencing the behavior of listeners via voice signals may be simply based on instrumental learning. However, voice signals are often produced strategically, and the senders' anticipation of the perceptual effects in listeners can help to produce appropriate (i.e., maximally efficient) voice signals, such as voice signals produced in a mating context (Charlton and Reby, 2016; Fraccaro et al., 2011; Pisanski et al., 2016; Wilkins et al., 2013) or to repulse predators (Charlton and Reby, 2016).

### 5.3. *Differential neural connectivity for voice signal production and perception*

In spite of the similarity of the neural systems involved, voice production and perception show a differential (intra- and inter-system) connectivity architecture to support their efficient functioning.

During voice signal production, the paramotor system shows some differential mapping with the PFC as origin or target. During voluntary voice motor planning, the PFC is the origin of an anterior-to-posterior mapping of voice programming, with M1 as the target that finally triggers voluntary voice motor movements (Hage and Nieder, 2016; Petkov and Jarvis, 2012). During voice motor execution, M1 maps motor commands back in a posterior-to-anterior mapping to the PMC and PFC which store an efference copy of the motor commands for a comparison against somatosensory and auditory feedback (Houde and Chang, 2015). A efference copy seems also mapped to the IPL (Bornkessel-Schlesewsky et al., 2015; Rauschecker and Scott, 2009), which might more directly compare actual vocal behavior against vocal plans using auditory and somatosensory feedback. The PMC, and probably also the PFC, is also the paramotor interface to the BG system involved in motor planning and voice learning (Bolhuis et al., 2010; Petkov and Jarvis, 2012; Ziegler and Ackermann, 2017). The limbic system is rather loosely connected with the paramotor system during voice production, but can receive information from the paramotor system during specific vocalizations that require planning and temporal sequencing of vocal behavioral elements (Frühholz et al., 2014a; Pichon and Kell, 2013).

As the central neural system for voice perception, the auditory system and specifically subregions of the auditory cortex receives and analyzes voice signals, and this information is directly or indirectly fed forward in a bottom-up manner to the paramotor system, including motor, premotor and primary sensorimotor regions, as target (Frühholz et al., 2015a; Rauschecker and Scott, 2009). Anterior subregions of paramotor system, such as the PFC, in turn, is the source of top-down predictions that influence both low- and high-level auditory regions (Friston and Frith, 2015; Smiley and Falchier, 2009). The paramotor system thus can figure both at the endpoint of bottom-up analyses of voice signal information, but it also can be the starting point for top-down modulations influencing voice signal analysis in the auditory system.

While perceiving voice signals, the higher-level auditory system in the aST and pST is also the link to several additional neural systems. First, the ST is the link to structures of the (para-)limbic system, such as the TP (Amaral and Price, 1984; Olson et al., 2007) (see above) and the amygdala (Frühholz et al., 2015b), that serves both the evaluation of the socio-affective meaning of sound signals (Kumar et al., 2012) and the retrieval of (episodic) memory associations (Frühholz et al., 2014b). These memory associations can facilitate decoding of ambiguous or corrupted voice signals. Second, the ST also serves auditory working memory (Pasternak and Greenlee, 2005; Scott et al., 2014) during extended or contextual voice signal perception in connection with the paramotor system and the hippocampus (Bolhuis and Gahr, 2006; Munoz-Lopez et al., 2010). Third, the ST, especially its posterior part, is also the brain region connecting to the BG (Ethofer et al., 2012; Frühholz and Grandjean, 2012; Péron et al., 2015). This altogether points to the notion the primary neural system for voice signal production and perception, respectively, is also the major neural interface to the additional neural system.

## 6. Songbird model of nonverbal auditory communication

The previous paragraphs mainly focused on the neural mechanisms of voice signal production and voice signal perception in primate species. However, some birds (especially parrots, humming birds, and songbirds) share some aspects of nonverbal voice communication with human primates (Jarvis et al., 2005). Unlike nonhuman primates, songbirds also share the vocal learning capabilities with humans (Bolhuis et al., 2010), although being less comparable in terms of the acoustic nature of vocal signaling. Neural mechanisms of song *production* in songbirds (Bolhuis et al., 2010; Bolhuis and Gahr, 2006; Bolhuis and Moorman, 2015) involves a similar neural architecture for voluntary and spontaneous voice signaling as in (human) primates.

First, neural mechanisms of song *production* in songbirds (Fig. 5, left panel) involves a similar neural architecture for voluntary and spontaneous voice signaling as in (human) primates. Song production is supported by similar four neural systems, such as paramotor (HVC, LMAN), limbic (RA), auditory (field L, CMM, NCM, NIf/Av), and BG circuit (Area X). The paramotor and limbic system represent the "vocal motor pathway" (VMP, green; sometimes also referred to as "song motor pathway", SMP) in the avian brain responsible for song initiation and sequencing (Solis et al., 2000). Motor song production in songbirds is centered on the HVC with connections to the limbic system (RA) figuring as the "vocal motor pathway" (VMP, green; sometimes also referred to as "song motor pathway", SMP). The HVC also interconnects with homologue regions of the striatum (X), the cortical auditory system (Field L, Av, NIf), midbrain structures (DM, a human PAG homolog), and brainstem motor nuclei (Xllts). The BG system and circuit is referred to as the "anterior forebrain pathway" (AFP, red) in the avian brain, and is also found in the brains of humans (but not yet in nonhuman primates (Ziegler and Ackermann, 2017)), with similar functional roles for vocal learning (Bolhuis and Gahr, 2006; Bolhuis and Moorman, 2015). The subcortical BG system interfaces with the cortical-like area LMAN. Both LMAN and the HVC might be homologue regions to Broca's area in human primates (Bolhuis et al., 2010), but this needs consistent confirmation by future studies. Lesion in the AFP during the song learning phase leads to premature vocal crystallizations (Ziegler and Ackermann, 2017), prevent plasticity of learned vocalization (Brainard and Doupe, 2000b), compromise efficient vocalizations monitoring in the vocal learning phase of songbirds (Prather et al., 2008), and seems to impair real-time evaluation of own vocalizations (Bolhuis et al., 2010). Song production also depends on feedback from own vocalizations and the perception of tutor vocalizations, which is mediated by the auditory systems including the NIf and CLM that feed auditory information to the HVC (Bauer et al., 2008).

Second, neural mechanisms of auditory and song *perception* in songbirds (Fig. 5, right panel) in the ascending auditory pathway connects the CN to cortex-like auditory regions (Field L, NIf, AV, CMM, NCM), which project to the HVC. Descending pathways modulate incoming auditory information via the HVC-RA-MLd axis. Thus, similar neural regions seem to be involved in song production and perception, although they show a differential neural connectivity. Third, the avian brain also seems to include neural circuits for predictive coding (Friston and Frith, 2015) during own song production and probably also during other songs perception, including predictions originating in the HVC and

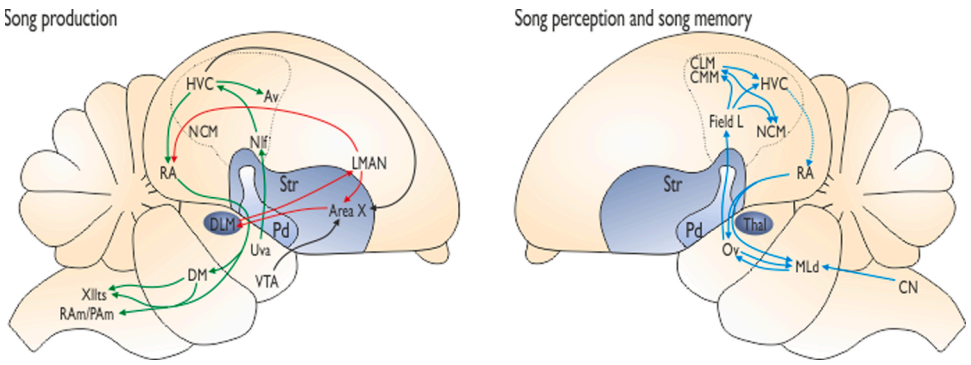


**Fig. 5.** Neural mechanisms of voice signal production and perception in songbirds.
Their neural mechanisms for song production and perception share many similarities with human primates. Song production (left panel) is supported by similar four neural systems, such as paramotor (HVC, LMAN), limbic (RA), auditory (field L, CMM, NCM, NIf/Av), and BG system and circuit (Area X). Song perception (right panel) in songbirds comprises subregions of the auditory (CLM, CMM, NCM, Field L), paramotor (HVC), and limbic system (RA).
Abbreviations: *Av* avalanche nucleus, *CMM* caudomedial mesopallium, *CN* cochlear nucleus, *DM* dorsomedial nucleus of the thalamus, HVC, *LMAN* lateral magnocellular nucleus of the anterior nidopallium, *NCM* caudomedial nidopallium, *NIf* nucleus interfacialis, *Ov* ovoidalis nucleus, *Pd* pallidum, *RA* robust nucleus of the arcopallium, *RF* reticular formation, *Uva* nucleus uvaeformis, *VTA* ventral tegmental area, *X* Area X of striatum, *XIIts* tracheosyringeal portion of the nucleus hypoglossus.

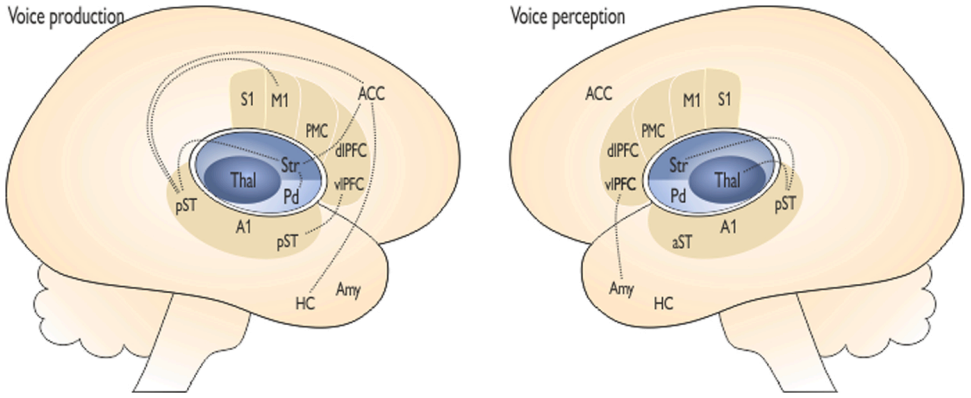


**Fig. 6.** Functional brain connectivity during voice signal production and perception.
Besides the rather established brain network underlying voice signal production and voice feedback processing (Fig. 1A), recent neuroimaging studies (Frühholz et al., 2015b, 2015c; Klaas et al., 2015; Pichon and Kell, 2013) in humans also pointed to an additional functional connectivity between brain regions (dashed black line), as identified largely by functional co-activations and temporal dynamics of brain signals (left panel). Although this connectivity may be based on direct structural connections between brain regions, note that these regions also could be functionally connected via intermediate nodes. Additionally, the direction of connections is usually not identified. Neuroimaging studies (Ethofer et al., 2012; Frühholz and Grandjean, 2012; Péron et al., 2015) on voice signal perception (right panel) also identified additional functional connectivity between regions, which go beyond the more established brain network models for voice signal perception. In addition to current evidence for these functional connections, future anatomical studies will need to delineate their structural bases.

Area X sent to the thalamus and the HVC, respectively. The neural mechanisms for voice signal production and perception in human primates thus seem paradoxically more similar in the avian brain than in the nonhuman primate brain, pointing to some nonlinear evolutionary lineage of evolved neural functions for nonverbal auditory communication (Bolhuis and Wynne, 2009).

## 7. Temporal dynamics of nonverbal auditory communication

The speed of primate voice signal production and comprehension is clearly remarkable (Fig. 3). Although typical natural speech may be slower, humans can produce and understand as many as 5–8 words per second. Despite the critical role of speed for speech comprehension in particular (Tallal, 2004), auditory voice signal perception also seems to be remarkably fast. For instance, although performance to recognize famous voices among unfamiliar distractors approaches an optimum only after at least 1.5-2 s of exposure, psychophysical studies suggest above-chance speaker identity recognition already with speech samples longer than 250 ms in duration (Schweinberger et al., 1997). In similar designs, above-chance vocal emotion recognition has been found even with durations as short as 180 ms (Pell and Kotz, 2011). Event-related brain potential studies have also identified remarkably early (150−200 ms) brain responses that appear to represent rapid discrimination of vocal versus nonvocal sounds (Charest et al., 2009), as well as responses to repetitions of voice characteristics (Sehweinberger, 2001; Zäske et al., 2009).

Evidence from EEG and MEG suggests particularly rapid processing of nonverbal emotional signals from the voice before 200 ms after voice onset (Jessen and Kotz, 2011a), and possibly before 100 ms for anger and fear signals (Jessen and Kotz, 2011b). By contrast, the processes mediating individual recognition of familiar speakers from the voice appear to onset in the region of 250−350 ms (Zäske et al., 2014). Although the interplay between production and perception is crucial during communication, research on the coupling between production

and perception is still sparse. Animal research has addressed avian duetting (Hall, 2004) and duetting in marmoset (i.e. antiphonal calling), which was accompanied by broad cFos gene expression in several frontal cortex areas compared to isolated vocal perception or production conditions (Miller et al., 2010), Furthermore, monkey research suggests a role of neurons in the frontal cortex, including a homolog of Broca's area, in vocal planning and initiation (Hage and Nieder, 2013), and possibly also during natural vocal behavior (Miller et al., 2015) when neuron activity appears to be coupled with vocalization pulses. In humans, initial evidence from MEG research suggests an online coupling between a reader's voice and a listener's cortical activity, particularly in the right pST (Bourguignon et al., 2013).

## 8. Contextual and multimodal aspects of vocal communication

Some environmental conditions, such as noise, do not modulate the voice signals itself, but rather change a sender's production of that signal. An important example for this context-dependent change of voice signal production is the "Lombard effect", such that vocal effort (i.e. intensity, vocal pitch, vocalization rate etc.) in speakers involuntarily increases in loud and noisy environment for compensation purposes accompanied by neural effects in the auditory cortex (Eliades and Wang, 2012). Moreover, multi-speaker environments also impose challenges for decoding information from a single target voice. Senders often adapt their voice to compensate for incriminatory effects of certain environments and transmission media (Tuomainen and Hazan, 2016). Finally, the perception of one type of nonverbal voice signal can be impaired when another signal is more dominant. For example, voice identity is less well identified in laughing voices than in neutral voices (Lavan et al., 2018).

Spatial distance between sender and listener is another contextual factor that influences voice signal production and perception. For example, cooperative vocal control in marmosets is mediated by vocal feedback such that vocalization intensity is adjusted to the spatial distance of listeners (Choi et al., 2015). Perceptual representations of space may deviate from the physical reality, such that vocal signals of threat may appear closer or more distant than they really are (Ceravolo et al., 2016). This perceptual representation and distortion of space based on the social importance of voice signals is encoded in the higher-level auditory system of listeners (Ceravolo et al., 2016). Overall, both voice production and perception are influenced by contextual conditions in which communication takes place.

Another important consideration refers to the fact that almost every vocal behavior is inevitably accompanied by facial and bodily motion, and this additional information can facilitate the perception of voice signal information. During voice production in noise, nonhuman primates, for example, combine vocal and facial information to enhance the detection of vocalizations (Chandrasekaran et al., 2011). The brain accordingly has evolved efficient mechanisms for rapid on-line processing of dynamic time-coupled communication signals from multiple channels (Belin et al., 2013; Ghazanfar, 2008; Schweinberger and Robertson, 2017; Sliwa et al., 2011; Yovel and O'Toole, 2016). Multimodal integration is accomplished by association areas of the auditory system (Anzellotti and Caramazza, 2017; Ghazanfar et al., 2008, 2005; Jessen et al., 2012; Perrodin et al., 2015) that support the recognition of physical attributes (Perrodin et al., 2015) or social voice information (Jessen et al., 2012). Although audiovisual integration may be most relevant here, vocal signals sometimes may also be accompanied by somatosensory (e.g. affective touch (Schirmer and Adolphs, 2017)) signals. Even olfactory signals may be relevant, at least with respect to the communication of more long-lasting social signals such as attractiveness (Groyecka et al., 2017).

More rapid multimodal integration of dynamic social signals in humans can be captured with methods that provide temporal resolution in the millisecond range. Electrophysiological studies suggest that bimodal (voice and face) stimuli trigger very rapid (<100 ms) audiovisual processing (Latinus et al., 2010; Schweinberger et al., 2011; Young, 2016), although audiovisual integration of complex cues that signal speaker identity occurs much later (>250 ms) and across sustained periods of time (Schweinberger et al., 2011). These studies underline that, like speech perception, speaker perception and nonverbal communication are inherently multimodal, and exploit efficient brain mechanisms for on-line processing of time-synchronized signals from different modalities (e.g., face, voice, body) (Belin et al., 2013; Schweinberger and Robertson, 2017; Yovel and O'Toole, 2016).

Face-voice integration in the perception of speaker identity likely involves direct structural connections from ventral temporal (fusiform) face areas to voice-sensitive areas in the aST (Blank et al., 2011), predominantly in the right hemisphere. Of note, face-voice integration appears to be particularly important for the recognition of well-known speakers (for which the brain has acquired specific crossmodal stimulus correspondences), whereas analogous effects are much smaller and more temporally limited in the case of unfamiliar speakers (González et al., 2011; Maguinness et al., 2018; Schweinberger and Robertson, 2017).

Audiovisual integration of emotional signals from the face and voice was already found in human infants aged seven months or less (Grossmann et al., 2006). In the adult human brain, audiovisual integration of emotional signals takes place even more rapidly than for identity signals. Audiovisual integration of emotional signals was observed within the first 200 ms or less (Jessen and Kotz, 2011b; Kokinous et al., 2015), and involves areas in the right ST(Awwad Shiekh Hasan et al., 2016; Hagan et al., 2013; Young, 2016), both in pST (Kreifelts et al., 2007; Perrodin et al., 2015; Watson et al., 2014) and more aST regions (Gainotti et al., 2008; Perrodin et al., 2015), and the amygdala (Milesi et al., 2014). Emotional nonverbal voice signal information specifically can influence the way we visually scan emotional faces (Rigoulot and Pell, 2014), but also the way we auditorily perceive verbal emotional information (Schirmer et al., 2004). Integrating vocal and facial signal information also concerns more complex socio-affective information, such as attractiveness (Groyecka et al., 2017), dominance and trustworthiness (Mileva et al., 2018).

Finally, although the precise time-course of face-voice integration has been mainly studied with respect to the perception of emotion and identity, there is beginning evidence that the formation of impressions about unfamiliar people (e.g., about ethnic background, competence, dominance, or trustworthiness) can also be influenced by multisensory signals (Mileva et al., 2018; Rakić et al., 2011a, 2011b).

## 9. Conclusions

Nonverbal auditory communication is a powerful way of exchanging socially relevant information. In humans, nonverbal voice signals convey social information that strongly influences social interactions, even when this influence may be implicit and poorly understood. This nonverbal voice channel in humans is shared with many other species and is used for both conspecific and heterospecific communication.

As nonverbal voice signals likely have evolved for the purpose of communication, they only seem relevant if we take at least a dyadic, i.e. pairwise, communicative perspective between a sender and a listener. Nonverbal voice production is only successful to the extent that voice signals are accurately decoded and have the intended effects in listeners. The evolution of nonverbal voice signals thus may exhibit co-dependency between effective voice signal production in listeners and accurate voice signal perception in listeners. This co-dependency on the behavioral vocal level is reflected in the similarity of the neural systems underlying voice signal production and perception.

In this review, we outlined three major neural systems that, although to different degrees, are involved in various functions underlying both voice signal production and perception. Not all three neural systems necessarily contribute to every instance of voice signal production and perception of various voice signal types, and not all these neural systems

are relevant for each vocally communicating species. However, those species that have an evolved repertoire of nonverbal voice signals and that are capable of voice signal learning, such as human primates, differentially recruit these three neural systems during the production and perception of physical attributes, social information, and non-arbitrary referential information in vocal communications. The similarity in the neural systems mediating voice production and perception and the differential connectivity between these systems both may have promoted the evolution of effective nonverbal communication.

## Author contributions

SF researched the article. SF and SRS contributed to discussions of the content and structure. SF and SRS contributed to writing the article and to reviewing and editing the manuscript before submission.

## Declaration of Competing Interest

The authors declare to have no competing interests.

## Acknowledgments

SF was supported by the Swiss National Science Foundation (SNSF PP00P1_157409/1 and PP00P1_183711/1).

## Appendix A. The Peer Review Overview and Supplementary data

The Peer Review Overview and Supplementary data associated with this article can be found in the online version, at doi:https://doi.org/10.1016/j.pneurobio.2019.101721.

## References

Abrams, D.A., Chen, T., Odriozola, P., Cheng, K.M., Baker, A.E., Padmanabhan, A., Ryali, S., Kochalka, J., Feinstein, C., Menon, V., 2016. Neural circuits underlying mother's voice perception predict social communication abilities in children. Proc. Natl. Acad. Sci. U. S. A. 113, 6295–6300. https://doi.org/10.1073/pnas.1602948113.

Ackermann, H., Hage, S.R., Ziegler, W., 2014. Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. Behav. Brain Sci. 72, 1–84. https://doi.org/10.1017/S0140525X13003099.

Agamaite, J.A., Chang, C.-J., Osmanski, M.S., Wang, X., 2015. A quantitative acoustic analysis of the vocal repertoire of the common marmoset (Callithrix jacchus). J. Acoust. Soc. Am. 138, 2906–2928. https://doi.org/10.1121/1.4934268.

Alba-Ferrara, L., Hausmann, M., Mitchell, R.L., Weis, S., 2011. The neural correlates of emotional prosody comprehension: disentangling simple from complex emotion. PLoS One 6, e28701. https://doi.org/10.1371/journal.pone.0028701.

Alderson-Day, B., Weis, S., McCarthy-Jones, S., Moseley, P., Smailes, D., Fernyhough, C., 2015. The brain's conversation with itself: neural substrates of dialogic inner speech. Soc. Cogn. Affect. Neurosci. 11, 110–120. https://doi.org/10.1093/scan/nsv094.

Amaral, D.G., Price, J.L., 1984. Amygdalo-cortical projections in the monkey (Macaca fascicularis). J. Comp. Neurol. 230, 465–496. https://doi.org/10.1002/cne.902300402.

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7, 268–277. https://doi.org/10.1038/nrn1884.

Andics, A., Gácsi, M., Faragó, T., Kis, A., Miklósi, Á., 2014. Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Curr. Biol. 24, 574–578. https://doi.org/10.1016/j.cub.2014.01.058.

Anikin, A., Lima, C.F., 2018. Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. Q. J. Exp. Psychol. 71, 622–641. https://doi.org/10.1080/17470218.2016.1270976.

Anikin, A., Bååth, R., Persson, T., 2018. Human non-linguistic vocal repertoire: call types and their meaning. J. Nonverbal Behav. 42, 53–80. https://doi.org/10.1007/s10919-017-0267-y.

Anolli, L., Ciceri, R., 1997. The voice of deception: vocal strategies of naive and able liars. J. Nonverbal Behav. 21, 259–284. https://doi.org/10.1023/A:1024916214403.

Anzellotti, S., Caramazza, A., 2017. Multimodal representations of person identity individuated with fMRI. Cortex 89, 85–97. https://doi.org/10.1016/j.cortex.2017.01.013.

Argyle, M., 1972. Non-verbal communication in human social interaction. Non-Verbal Communication. Cambridge U. Press, Oxford, England, p. 443 p. xiii.

Arnal, L.H., Flinker, A., Kleinschmidt, A., Giraud, A.L., Poeppel, D., 2015. Human screams occupy a privileged niche in the communication soundscape. Curr. Biol. 25, 2051–2056. https://doi.org/10.1016/j.cub.2015.06.043.

Arnold, C., Gehrig, J., Gispert, S., Seifried, C., Kell, C.A., 2014. Pathomechanisms and compensatory efforts related to Parkinsonian speech. Neuroimage Clin. 4, 82–97. https://doi.org/10.1016/j.nicl.2013.10.016.

Ashby, F.G., Turner, B.O., Horvitz, J.C., 2010. Cortical and basal ganglia contributions to habit learning and automaticity. Trends Cogn. Sci. 14, 208–215. https://doi.org/10.1016/j.tics.2010.02.001.

Averbeck, B.B., Romanski, L.M., 2004. Principal and independent components of macaque vocalizations: constructing stimuli to probe high-level sensory processing. J. Neurophysiol. 91, 2897–2909. https://doi.org/10.1152/jn.01103.2003.

Awwad Shiekh Hasan, B., Valdes-Sosa, M., Gross, J., Belin, P., 2016. "Hearing faces and seeing voices": amodal coding of person identity in the human brain. Sci. Rep. 6 https://doi.org/10.1038/srep37494.

Babel, M., McGuire, G., King, J., 2014. Towards a more nuanced view of vocal attractiveness. PLoS One 9, e88616. https://doi.org/10.1371/journal.pone.0088616.

Bachorowski, J.A., Owren, M.J., 2003. Sounds of emotion: production and perception of affect-related vocal acoustics. Annals of the New York Academy of Sciences, pp. 244–265. https://doi.org/10.1196/annals.1280.012.

Baltaxe, C.A.M., Simmons, J.Q., 1985. Prosodic development in Normal and autistic children. In: Schopler, E., Mesibov, G.B. (Eds.), Communication Problems in Autism. Springer US, Boston, MA, pp. 95–125. https://doi.org/10.1007/978-1-4757-4806-2_7.

Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. J. Pers. Soc. Psychol. 70, 614–636. https://doi.org/10.1037/0022-3514.70.3.614.

Barsalou, L.W., 2009. Simulation, situated conceptualization, and prediction. Philos. Trans. R. Soc. B Biol. Sci. 364, 1281–1289. https://doi.org/10.1098/rstb.2008.0319.

Bauer, E.E., Coleman, M.J., Roberts, T.F., Roy, A., Prather, J.F., Mooney, R., 2008. A synaptic basis for auditory-vocal integration in the songbird. J. Neurosci. 28, 1509–1522. https://doi.org/10.1523/JNEUROSCI.3838-07.2008.

Behroozmand, R., Karvelis, L., Liu, H., Larson, C.R., 2009. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. Clin. Neurophysiol. 120, 1303–1312. https://doi.org/10.1016/j.clinph.2009.04.022.

Belin, P., Zatorre, R.J., Lafallie, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312. https://doi.org/10.1038/35002078.

Belin, P., Fecteau, S., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135. https://doi.org/10.1016/j.tics.2004.01.008.

Belin, P., Campanella, S., Ethofer, T., 2013. Integrating face and voice in person perception. Integrating Face and Voice in Person Perception. https://doi.org/10.1007/978-1-4614-3585-3.

Belin, P., Boehme, B., McAleer, P., 2017. The sound of trustworthiness: acoustic-based modulation of perceived voice personality. PLoS One 12, 1–9. https://doi.org/10.1371/journal.pone.0185651.

Belin, P., Bodin, C., Aglieri, V., 2018. A "voice patch" system in the primate brain for processing vocal information? Hear. Res. 366, 65–74. https://doi.org/10.1016/j.heares.2018.04.010.

Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14, 2105–2109. https://doi.org/10.1097/00001756-200311140-00019.

Benetti, S., Van Ackeren, M.J., Rabini, G., Zonca, J., Foa, V., Baruffaldi, F., Rezk, M., Pavani, F., Rossion, B., Collignon, O., 2017. Functional selectivity for face processing in the temporal voice area of early deaf individuals. Proc. Natl. Acad. Sci. U. S. A. 114, E6437–E6446. https://doi.org/10.1073/pnas.1618287114.

Bestelmeyer, P.E.G., Latinus, M., Bruckert, L., Rouger, J., Crabbe, F., Belin, P., 2012. Implicitly perceived vocal attractiveness modulates prefrontal cortex activity. Cereb. Cortex 22, 1263–1270. https://doi.org/10.1093/cercor/bhr204.

Bestelmeyer, P.E.G., Maurage, P., Rouger, J., Latinus, M., Belin, P., 2014. Adaptation to vocal expressions reveals multistep perception of auditory emotion. J. Neurosci. 34, 8098–8105. https://doi.org/10.1523/JNEUROSCI.4820-13.2014.

Bestelmeyer, P.E.G., Belin, P., Ladd, D.R., 2015. A neural marker for social bias toward in-group accents. Cereb. Cortex 25, 3953–3961. https://doi.org/10.1093/cercor/bhu282.

Blank, H., Anwander, A., von Kriegstein, K., 2011. Direct structural connections between voice-and face-recognition areas. J. Neurosci. 31, 12906–12915. https://doi.org/10.1523/JNEUROSCI.2091-11.2011.

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G.J., Renvall, V., Deoni, S., Gasston, D., Williams, S.C.R., Johnson, M.H., Simmons, A., Murphy, D.G.M., 2011. Early specialization for voice and emotion processing in the infant brain. Curr. Biol. 21, 1220–1224. https://doi.org/10.1016/j.cub.2011.06.009.

Bolhuis, J.J., Gahr, M., 2006. Neural mechanisms of birdsong memory. Nat. Rev. Neurosci. 7, 347–357. https://doi.org/10.1038/nrn1904.

Bolhuis, J.J., Moorman, S., 2015. Birdsong memory and the brain: in search of the template. Neurosci. Biobehav. Rev. https://doi.org/10.1016/j.neubiorev.2014.11.019.

Bolhuis, J.J., Wynne, C.D.L., 2009. Can evolution explain how minds work? Nature 458, 832–833. https://doi.org/10.1038/458832a.

Bolhuis, J.J., Okanoya, K., Scharff, C., 2010. Twitter evolution: converging mechanisms in birdsong and human speech. Nat. Rev. Neurosci. 11, 747–759. https://doi.org/10.1038/nrn2931.

Borkowska, B., Pawlowski, B., 2011. Female voice frequency in the context of dominance and attractiveness perception. Anim. Behav. 82, 55–59. https://doi.org/10.1016/j.anbehav.2011.03.024.

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S.L., Rauschecker, J.P., 2015. Neurobiological roots of language in primate audition: common computational

properties. Trends Cogn. Sci. 19, 142–150. https://doi.org/10.1016/j.tics.2014.12.008.
Bourguignon, M., De Tiège, X., De Beeck, M.O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., Jousmäki, V., 2013. The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Hum. Brain Mapp. 34, 314–326. https://doi.org/10.1002/hbm.21442.
Brainard, M.S., Doupe, A.J., 2000a. Auditory feedback in learning and maintenance of vocal behaviour. Nat. Rev. Neurosci. 1, 31–40. https://doi.org/10.1038/35036205.
Brainard, M.S., Doupe, A.J., 2000b. Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. Nature 404, 762–766. https://doi.org/10.1038/35008083.
Brosch, T., Grandjean, D., Sander, D., Scherer, K.R., 2009. Cross-modal emotional attention: emotional voices modulate early stages of visual processing. J. Cogn. Neurosci. 21, 1670–1679. https://doi.org/10.1162/jocn.2009.21110.
Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G.A., Kawahara, H., Belin, P., 2010. Vocal attractiveness increases by averaging. Curr. Biol. 20, 116–120. https://doi.org/10.1016/j.cub.2009.11.034.
Caekebeke, J.F.V., Jennekens-Schinkel, A., Van der Linden, M.E., Buruma, O.J.S., Roos, R.A.C., 1991. The interpretation of dysprosody in patients with Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 54, 145–148. https://doi.org/10.1136/jnnp.54.2.145.
Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C.R., McGuire, P. K., Woodruff, P.W.R., Iversen, S.D., David, A.S., 1997. Activation of auditory cortex during silent lipreading. Science 276 (80-), 593–596. https://doi.org/10.1126/science.276.5312.593.
Capilla, A., Belin, P., Gross, J., 2013. The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. Cereb. Cortex 23, 1388–1395. https://doi.org/10.1093/cercor/bhs119.
Cäsar, C., Zuberbühler, K., Young, R.J., Byrne, R.W., 2013. Titi monkey call sequences vary with predator location and type. Biol. Lett. 9 https://doi.org/10.1098/rsbl.2013.0535.
Ceravolo, L., Frühholz, S., Grandjean, D., 2016. Modulation of auditory spatial attention by angry prosody: an fMRI auditory dot-probe study. Front. Neurosci. 10 (May) https://doi.org/10.3389/fnins.2016.00216.
Chandrasekaran, C., Lemus, L., Trubanova, A., Gondan, M., Ghazanfar, A.A., 2011. Monkeys and humans share a common computation for face/voice integration. PLoS Comput. Biol. 7 https://doi.org/10.1371/journal.pcbi.1002165.
Charest, I., Pernet, C.R., Rousselet, G.A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., Chartrand, J.P., Belin, P., 2009. Electrophysiological evidence for an early processing of human voices. BMC Neurosci. 10, 127. https://doi.org/10.1186/1471-2202-10-127.
Charlton, B.D., Reby, D., 2016. The evolution of acoustic size exaggeration in terrestrial mammals. Nat. Commun. 7 https://doi.org/10.1038/ncomms12739.
Cheang, H.S., Pell, M.D., 2008. The sound of sarcasm. Speech Commun. 50, 366–381. https://doi.org/10.1016/j.specom.2007.11.003.
Cheung, S.W., Nagarajan, S.S., Schreiner, C.E., Bedenbaugh, P.H., Wong, A., 2005. Plasticity in primary auditory cortex of monkeys with altered vocal production. J. Neurosci. 25, 2490–2503. https://doi.org/10.1523/JNEUROSCI.5289-04.2005.
Cheung, C., Hamiton, L.S., Johnson, K., Chang, E.F., 2016. The auditory representation of speech sounds in human motor cortex. Elife 5. https://doi.org/10.7554/eLife.12577.
Choi, J.Y., Takahashi, D.Y., Ghazanfar, A.A., 2015. Cooperative vocal control in marmoset monkeys via vocal feedback. J. Neurophysiol. 114, 274–283. https://doi.org/10.1152/jn.00228.2015.
Clarke, E., Reichard, U.H., Zuberbühler, K., 2015. Context-specific close-range "hoo" calls in wild gibbons (Hylobates lar). BMC Evol. Biol. 15, 56. https://doi.org/10.1186/s12862-015-0332-2.
Constantinidis, C., Procyk, E., 2004. The primate working memory networks. Cogn. Affect. Behav. Neurosci. 4, 444–465. https://doi.org/10.3758/CABN.4.4.444.
Creutzfeldt, O., Ojemann, G., Lettich, E., 1989. Neuronal activity in the human lateral temporal lobe - I. Responses to speech. Exp. Brain Res. 77, 451–475. https://doi.org/10.1007/BF00249600.
D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., Fadiga, L., 2009. The motor somatotopy of speech perception. Curr. Biol. 19, 381–385. https://doi.org/10.1016/j.cub.2009.01.017.
Dawkins, R., Krebs, J.R., 1978. Animal signals: information or manipulation? In: Krebs, J.R., Davies, N.B. (Eds.), Behavioural Ecology: An Evolutionary Approach. Sinauer Associates, Inc., Sutherland, MA.
De La Torre, S., Snowdon, C.T., 2009. Dialects in pygmy marmosets? Population variation in call structure. Am. J. Primatol. 71, 333–342. https://doi.org/10.1002/ajp.20657.
De Lucia, M., Clarke, S., Murray, M.M., 2010. A temporal hierarchy for conspecific vocalization discrimination in humans. J. Neurosci. 30, 11210–11221. https://doi.org/10.1523/JNEUROSCI.2239-10.2010.
Doupe, A.J., Kuhl, P.K., 1999. Birdsong and human speech: common themes and mechanisms. Annu. Rev. Neurosci. 22, 567–631. https://doi.org/10.1146/annurev.neuro.22.1.567.
Dricu, M., Frühholz, S., 2016. Perceiving emotional expressions in others: activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. Neurosci. Biobehav. Rev. 71, 810–828. https://doi.org/10.1016/j.neubiorev.2016.10.020.
Dricu, M., Ceravolo, L., Grandjean, D., Frühholz, S., 2017. Biased and unbiased perceptual decision-making on vocal emotions. Sci. Rep. 7 https://doi.org/10.1038/s41598-017-16594-w.
Egnor, S.E.R., Hauser, M.D., 2004. A paradox in the evolution of primate vocal learning. Trends Neurosci. 27, 649–654. https://doi.org/10.1016/j.tins.2004.08.009.
Egnor, S.E.R., Iguina, C.G., Hauser, M.D., 2006. Perturbation of auditory feedback causes systematic perturbation in vocal structure in adult cotton-top tamarins. J. Exp. Biol. 209, 3652–3663. https://doi.org/10.1242/jeb.02420.
Ehret, G.G., 2006. Common rules of communication sound perception. In: Kanwal, J.S., Ehret, G. (Eds.), Behaviour and Neurodynamics for Auditory Communication. Cambridge University Press, Cambridge, pp. 85–114.
Eliades, S.J., Tsunada, J., 2018. Auditory cortical activity drives feedback-dependent vocal control in marmosets. Nat. Commun. 9 https://doi.org/10.1038/s41467-018-04961-8.
Eliades, S.J., Wang, X., 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453, 1102–1106. https://doi.org/10.1038/nature06910.
Eliades, S.J., Wang, X., 2012. Neural correlates of the lombard effect in primate auditory cortex. J. Neurosci. 32, 10737–10748. https://doi.org/10.1523/JNEUROSCI.3448-11.2012.
Ellis, H.D., Young, A.W., Quayle, A.H., De Pauw, K.W., 1997. Reduced autonomic responses to faces in Capgras delusion. Proc. R. Soc. B Biol. Sci. 264, 1085–1092. https://doi.org/10.1098/rspb.1997.0150.
Engelberg, J.W.M., Gouzoules, H., 2019. The credibility of acted screams: implications for emotional communication research. Q. J. Exp. Psychol. 72, 1889–1902. https://doi.org/10.1177/1747021818816307.
Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P., 2012. Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cereb. Cortex 22, 191–200. https://doi.org/10.1093/cercor/bhr113.
Fichtel, C., Perry, S., Gros-Louis, J., 2005. Alarm calls of white-faced capuchin monkeys: an acoustic analysis. Anim. Behav. 70, 165–176. https://doi.org/10.1016/j.anbehav.2004.09.020.
Fitch, W.T.S., 2002. Primate vocal production and its implications for auditory research. Primate Audit. Ethol. Neurobiol. 87–108. https://doi.org/10.1201/9781420041224.ch6.
Ford, J.M., Roach, B.J., Faustman, W.O., Mathalon, D.H., 2007. Synch before you speak: auditory hallucinations in schizophrenia. Am. J. Psychiatry 164, 458–466. https://doi.org/10.1176/ajp.2007.164.3.458.
Ford, J.M., Roach, B.J., Jorgensen, K.W., Turner, J.A., Brown, G.G., Notestine, R., Bischoff-Grethe, A., Greve, D., Wible, C., Lauriello, J., Belger, A., Mueller, B.A., Calhoun, V., Preda, A., Keator, D., O'Leary, D.S., Lim, K.O., Glover, G., Potkin, S.G., Mathalon, D.H., 2009. Tuning in to the voices: a multisite fMRI study of auditory hallucinations. Schizophr. Bull. 35, 58–66. https://doi.org/10.1093/schbul/sbn140.
Fraccaro, P., Jones, B., Vukovic, J., Smith, F., Watkins, C., Feinberg, D., Little, A., Debruine, L., 2011. Experimental evidence that women speak in a higher voice pitch to men they find attractive. J. Evol. Psychol. 9, 57–67. https://doi.org/10.1556/JEP.9.2011.33.1.
Fraccaro, P.J., O'Connor, J.J.M., Re, D.E., Jones, B.C., DeBruine, L.M., Feinberg, D.R., 2013. Faking it: deliberately altered voice pitch and vocal attractiveness. Anim. Behav. 85, 127–136. https://doi.org/10.1016/j.anbehav.2012.10.016.
Friederici, A.D., 2011. The brain basis of language processing: from structure to function. Physiol. Rev. 91, 1357–1392. https://doi.org/10.1152/physrev.00006.2011.
Friederici, A.D., 2012. The cortical language circuit: from auditory perception to sentence comprehension. Trends Cogn. Sci. 16, 262–268. https://doi.org/10.1016/j.tics.2012.04.001.
Friston, K.J., Frith, C.D., 2015. Active inference, Communication and hermeneutics. Cortex 68, 129–143. https://doi.org/10.1016/j.cortex.2015.03.025.
Frühholz, S., Grandjean, D., 2012. Towards a fronto-temporal neural network for the decoding of angry vocal expressions. Neuroimage 62, 1658–1666. https://doi.org/10.1016/j.neuroimage.2012.06.015.
Frühholz, S., Grandjean, D., 2013a. Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. Neurosci. Biobehav. Rev. 37, 24–35. https://doi.org/10.1016/j.neubiorev.2012.11.002.
Frühholz, S., Grandjean, D., 2013b. Amygdala subregions differentially respond and rapidly adapt to threatening voices. Cortex 49, 1394–1403. https://doi.org/10.1016/j.cortex.2012.08.003.
Frühholz, S., Staib, M., 2017. Neurocircuitry of impaired affective sound processing: a clinical disorders perspective. Neurosci. Biobehav. Rev. 83, 516–524. https://doi.org/10.1016/j.neubiorev.2017.09.009.
Frühholz, S., Sander, D., Grandjean, D., 2014a. Functional neuroimaging of human vocalizations and affective speech. Behav. Brain Sci. 37, 554–555. https://doi.org/10.1017/S0140525X13004020.
Frühholz, S., Trost, W., Grandjean, D., 2014b. The role of the medial temporal limbic system in processing emotions in voice and music. Prog. Neurobiol. 123, 1–17. https://doi.org/10.1016/j.pneurobio.2014.09.003.
Frühholz, S., Gschwind, M., Grandjean, D., 2015a. Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. Neuroimage 109, 27–34. https://doi.org/10.1016/j.neuroimage.2015.01.016.
Frühholz, S., Hofstetter, C., Cristinzio, C., Saj, A., Seeck, M., Vuilleumier, P., Grandjean, D., 2015b. Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. Proc. Natl. Acad. Sci. U. S. A. 112, 1583–1588. https://doi.org/10.1073/pnas.1411315112.
Frühholz, S., Klaas, H.S., Patel, S., Grandjean, D., 2015c. Talking in fury: the cortico-subcortical network underlying angry vocalizations. Cereb. Cortex 25, 2752–2762. https://doi.org/10.1093/cercor/bhu074.
Frühholz, S., Trost, W., Grandjean, D., 2016a. Whispering - the hidden side of auditory communication. Neuroimage 142, 602–612. https://doi.org/10.1016/j.neuroimage.2016.08.023.

Frühholz, S., Trost, W., Kotz, S.A., 2016b. The sound of emotions-Towards a unifying neural network perspective of affective sound processing. Neurosci. Biobehav. Rev. 68, 1–15. https://doi.org/10.1016/j.neubiorev.2016.05.002.
Frühholz, S., van der Zwaag, W., Saenz, M., Belin, P., Schobert, A.K., Vuilleumier, P., Grandjean, D., 2016c. Neural decoding of discriminative auditory object features depends on their socio-affective valence. Soc. Cogn. Affect. Neurosci. 11, 1638–1649. https://doi.org/10.1093/scan/nsw066.
Frühholz, S., Ceravolo, L., Frühholz, S., Ceravolo, L., 2018. The neural network underlying the processing of affective vocalizations. The Oxford Handbook of Voice Perception. Oxford University Press, Oxford, UK, pp. 430–458. https://doi.org/10.1093/oxfordhb/9780198743187.013.19.
Gainotti, G., Ferraccioli, M., Quaranta, D., Marra, C., 2008. Cross-modal recognition disorders for persons and other unique entities in a patient with right fronto-temporal degeneration. Cortex 44, 238–248. https://doi.org/10.1016/j.cortex.2006.09.001.
Gazzola, V., Aziz-Zadeh, L., Keysers, C., 2006. Empathy and the somatotopic auditory mirror system in humans. Curr. Biol. 16, 1824–1829. https://doi.org/10.1016/j.cub.2006.07.072.
Geiser, E., Notter, M., Gabrieli, J.D.E., 2012. A corticostriatal neural system enhances auditory perception through temporal context processing. J. Neurosci. 32, 6177–6182. https://doi.org/10.1523/JNEUROSCI.5153-11.2012.
Ghazanfar, A.A., 2008. Language evolution: neural differences that make a difference. Nat. Neurosci. 11, 382–384. https://doi.org/10.1038/nn0408-382.
Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., Logothetis, N.K., 2005. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J. Neurosci. 25, 5004–5012. https://doi.org/10.1523/JNEUROSCI.0799-05.2005.
Ghazanfar, A.A., Turesson, H.K., Maier, J.X., van Dinther, R., Patterson, R.D., Logothetis, N.K., 2007. Vocal-tract resonances as indexical cues in Rhesus monkeys. Curr. Biol. 17, 425–430. https://doi.org/10.1016/j.cub.2007.01.029.
Ghazanfar, A.A., Chandrasekaran, C., Logothetis, N.K., 2008. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. J. Neurosci. 28, 4457–4469. https://doi.org/10.1523/JNEUROSCI.0541-08.2008.
Goldman, A.I., Sripada, C.S., 2005. Simulationist models of face-based emotion recognition. Cognition. https://doi.org/10.1016/j.cognition.2004.01.005.
González, I.Q., León, M.A.B., Belin, P., Martínez-Quintana, Y., García, L.G., Castillo, M.S., 2011. Person identification through faces and voices: an ERP study. Brain Res. 1407, 13–26. https://doi.org/10.1016/j.brainres.2011.03.029.
Grahn, J.A., Brett, M., 2007. Rhythm and beat perception in motor areas of the brain. J. Cogn. Neurosci. 19, 893–906. https://doi.org/10.1162/jocn.2007.19.5.893.
Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P., 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat. Neurosci. 8, 145–146. https://doi.org/10.1038/nn1392.
Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348–353. https://doi.org/10.1016/S0166-2236(02)02191-4.
Grossmann, T., Striano, T., Friederic, A.D., 2006. Crossmodal integration of emotional information from face and voice in the infant brain. Dev. Sci. 9, 309–315. https://doi.org/10.1111/j.1467-7687.2006.00494.x.
Grossmann, T., Oberecker, R., Koch, S.P., Friederici, A.D., 2010. The developmental origins of voice processing in the human brain. Neuron 65, 852–858. https://doi.org/10.1016/j.neuron.2010.03.001.
Groyecka, A., Pisanski, K., Sorokowska, A., Havlícek, J., Karwowski, M., Puts, D., Craig Roberts, S., Sorokowski, P., 2017. Attractiveness is multimodal: beauty is also in the nose and ear of the beholder. Front. Psychol. 8 https://doi.org/10.3389/fpsyg.2017.00778.
Hagan, C.C., Woods, W., Johnson, S., Green, G.G.R., Young, A.W., 2013. Involvement of right STS in audio-visual integration for affective speech demonstrated using MEG. PLoS One 8. https://doi.org/10.1371/journal.pone.0070648.
Hage, S.R., Nieder, A., 2013. Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. Nat. Commun. 4, 2409. https://doi.org/10.1038/ncomms3409.
Hage, S.R., Nieder, A., 2016. Dual neural network model for the evolution of speech and language. Trends Neurosci. 39, 813–829. https://doi.org/10.1016/j.tins.2016.10.006.
Hage, S.R., Gavrilov, N., Nieder, A., 2013. Cognitive control of distinct vocalizations in rhesus monkeys. J. Cogn. Neurosci. 25, 1692–1701. https://doi.org/10.1162/jocn_a_00428.
Hahnloser, R.H.R., Kotowicz, A., 2010. Auditory representations and memory in birdsong learning. Curr. Opin. Neurobiol. 20, 332–339. https://doi.org/10.1016/j.conb.2010.02.011.
Hall, M.L., 2004. A review of hypotheses for the functions of avian duetting. Behav. Ecol. Sociobiol. https://doi.org/10.1007/s00265-003-0741-x.
Harnsberger, J.D., Brown, W.S., Shrivastav, R., Rothman, H., 2010. Noise and tremor in the perception of vocal aging in males. J. Voice 24, 523–530. https://doi.org/10.1016/j.jvoice.2009.01.003.
Hass, J., Herrmann, J.M., 2012. The neural representation of time: an information-theoretic perspective. Neural Comput. 24, 1519–1552. https://doi.org/10.1162/NECO_a_00280.
Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. Neuroscience: The faculty of language: What is it, who has it, and how did it evolve? Science 298 (80-), 1569–1579. https://doi.org/10.1126/science.298.5598.1569.
Hellbernd, N., Sammler, D., 2016. Prosody conveys speaker's intentions: acoustic cues for speech act perception. J. Mem. Lang. 88, 70–86. https://doi.org/10.1016/j.jml.2016.01.001.
Hickok, G., 2012. Computational neuroanatomy of speech production. Nat. Rev. Neurosci. 13, 135–145. https://doi.org/10.1038/nrn3158.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402. https://doi.org/10.1038/nrn2113.
Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69, 407–422. https://doi.org/10.1016/j.neuron.2011.01.019.
Hollien, H., Geison, L., Hicks, J.W., 1987. Voice stress evaluators and lie detection. J. Forensic Sci. 32, 11143J https://doi.org/10.1520/jfs11143j.
Houde, J.F., Chang, E.F., 2015. The cortical computations underlying feedback control in vocal production. Curr. Opin. Neurobiol. 33, 174–181. https://doi.org/10.1016/j.conb.2015.04.006.
Hugdahl, K., 2009. "Hearing voices": auditory hallucinations as failure of top-down control of bottom-up perceptual processes. Scand. J. Psychol. 50, 553–560. https://doi.org/10.1111/j.1467-9450.2009.00775.x.
Hughes, S.M., Farley, S.D., Rhodes, B.C., 2010. Vocal and physiological changes in response to the physical attractiveness of conversational partners. J. Nonverbal Behav. 34, 155–167. https://doi.org/10.1007/s10919-010-0087-9.
Jarvis, E.D., 2007. Neural systems for vocal learning in birds and humans: a synopsis. J. Ornithol. 148 (SUPPL. 1) https://doi.org/10.1007/s10336-007-0243-0.
Jarvis, E., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D.J., Shimizu, T., Striedter, G., Martin Wild, J., Ball, G.F., Dugas-Ford, J., Durand, S.E., Hough, G.E., Husband, S., Kubikova, L., Lee, D.W., Mello, C. V., Powers, A., Siang, C., Smulders, T.V., Wada, K., White, S.A., Yamamoto, K., Yu, J., Reiner, A., Butler, A.B., 2005. Avian brains and a new understanding of vertebrate brain evolution. Nat. Rev. Neurosci. 6, 151–159. https://doi.org/10.1038/nrn1606.
Jasmin, K., Lima, C.F., Scott, S.K., 2019. Understanding rostral–caudal auditory cortex contributions to auditory perception. Nat. Rev. Neurosci. 20, 425–434. https://doi.org/10.1038/s41583-019-0160-2.
Jessen, S., Kotz, S.A., 2011a. The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. Neuroimage 58, 665–674. https://doi.org/10.1016/j.neuroimage.2011.06.035.
Jessen, S., Kotz, S.A., 2011b. The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. Neuroimage 58, 665–674. https://doi.org/10.1016/j.neuroimage.2011.06.035.
Jessen, S., Obleser, J., Kotz, S.A., 2012. How bodies and voices interact in early emotion perception. PLoS One 7, e36070. https://doi.org/10.1371/journal.pone.0036070.
Jiang, X., Pell, M.D., 2015. On how the brain decodes vocal cues about speaker confidence. Cortex 66, 9–34. https://doi.org/10.1016/j.cortex.2015.02.002.
Joosten, B., Postma, E., Krahmer, E., 2015. Voice activity detection based on facial movement. J. Multimodal User Interfaces 9, 183–193. https://doi.org/10.1007/s12193-015-0187-2.
Jürgens, U., 2002. Neural pathways underlying vocal control. Neurosci. Biobehav. Rev. 26, 235–258. https://doi.org/10.1016/S0149-7634(01)00068-9.
Jürgens, U., Ploog, D., 1970. Cerebral representation of vocalization in the squirrel monkey. Exp. Brain Res. 10, 532–554. https://doi.org/10.1007/BF00234269.
Kaplan, J.T., Aziz-Zadeh, L., Uddin, L.Q., Iacoboni, M., 2008. The self across the senses: an fMRI study of self-face and self-voice recognition. Soc. Cogn. Affect. Neurosci. 3, 218–223. https://doi.org/10.1093/scan/nsn014.
Kersken, V., Zuberbühler, K., Gomez, J.C., 2017. Listeners can extract meaning from non-linguistic infant vocalisations cross-culturally. Sci. Rep. 7, 41016. https://doi.org/10.1038/srep41016.
Keysers, C., Kaas, J.H., Gazzola, V., 2010. Somatosensation in social perception. Nat. Rev. Neurosci. 11, 417–428. https://doi.org/10.1038/nrn2833.
Klaas, H.S., Frühholz, S., Grandjean, D., 2015. Aggressive vocal expressions-an investigation of their underlying neural network. Front. Behav. Neurosci. 9, 121. https://doi.org/10.3389/fnbeh.2015.00121.
Ko, S.J., Sadler, M.S., Galinsky, A.D., 2015. The sound of power: conveying and detecting hierarchical rank through voice. Psychol. Sci. 26, 3–14. https://doi.org/10.1177/0956797614553009.
Kojima, S., Kao, M.H., Doupe, A.J., 2013. Task-related "cortical" bursting depends critically on basal ganglia input and is linked to vocal plasticity. Proc. Natl. Acad. Sci. U. S. A. 110, 4756–4761. https://doi.org/10.1073/pnas.1216308110.
Kokinous, J., Kotz, S.A., Tavano, A., Schröger, E., 2015. The role of emotion in dynamic audiovisual integration of faces and voices. Soc. Cogn. Affect. Neurosci. 10, 713–720. https://doi.org/10.1093/scan/nsu105.
Korb, S., Frühholz, S., Grandjean, D., 2014. Reappraising the voices of wrath. Soc. Cogn. Affect. Neurosci. 10, 1644–1660. https://doi.org/10.1093/scan/nsv051.
Korzyukov, O., Karvelis, L., Behroozmand, R., Larson, C.R., 2012. ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback. Int. J. Psychophysiol. 83, 71–78. https://doi.org/10.1016/j.ijpsycho.2011.10.006.
Kotz, S.A., Schwartze, M., 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. Trends Cogn. Sci. 14, 392–399. https://doi.org/10.1016/j.tics.2010.06.005.
Kotz, S.A., Schwartze, M., Schmidt-Kassow, M., 2009. Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. Cortex 45, 982–990. https://doi.org/10.1016/j.cortex.2009.02.010.
Kragel, P.A., LaBar, K.S., 2016. Somatosensory representations link the perception of emotional expressions and sensory experience. eNeuro 3, 169–177. https://doi.org/10.1523/ENEURO.0090-15.2016.
Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., Wildgruber, D., 2007. Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. Neuroimage 37, 1445–1456. https://doi.org/10.1016/j.neuroimage.2007.06.020.
Kroodsma, D.E., Pickert, R., 1984. Repertoire size, auditory templates, and selective vocal learning in songbirds. Anim. Behav. 32, 395–399. https://doi.org/10.1016/S0003-3472(84)80275-4.

Kumar, S., Stephan, K.E., Warren, J.D., Friston, K.J., Griffiths, T.D., 2007. Hierarchical processing of auditory objects in humans. PLoS Comput. Biol. 3, 0977–0985. https://doi.org/10.1371/journal.pcbi.0030100.
Kumar, S., von Kriegstein, K., Friston, K., Griffiths, T.D., 2012. Features versus feelings: dissociable representations of the acoustic features and valence of aversive sounds. J. Neurosci. 32, 14184–14192. https://doi.org/10.1523/JNEUROSCI.1759-12.2012.
Kumar, S., Joseph, S., Gander, P.E., Barascud, N., Halpern, A.R., Griffiths, T.D., 2016. A brain system for auditory working memory. J. Neurosci. 36, 4492–4505. https://doi.org/10.1523/JNEUROSCI.4341-14.2016.
Latinus, M., Belin, P., 2011. Anti-voice adaptation suggests prototype-based coding of voice identity. Front. Psychol. 2 https://doi.org/10.3389/fpsyg.2011.00175.
Latinus, M., VanRullen, R., Taylor, M.J., 2010. Top-down and bottom-up modulation in processing bimodal face/voice stimuli. BMC Neurosci. 11 https://doi.org/10.1186/1471-2202-11-36.
Latinus, M., McAleer, P., Bestelmeyer, P.E.G., Belin, P., 2013. Norm-based coding of voice identity in human auditory cortex. Curr. Biol. 23, 1075–1080. https://doi.org/10.1016/j.cub.2013.04.055.
Laukka, P., Åhs, F., Furmark, T., Fredrikson, M., 2011. Neurofunctional correlates of expressed vocal affect in social phobia. Cogn. Affect. Behav. Neurosci. 11, 413–425. https://doi.org/10.3758/s13415-011-0032-3.
Lauterbach, E.C., Cummings, J.L., Kuppuswamy, P.S., 2013. Toward a more precise, clinically-informed pathophysiology of pathological laughing and crying. Neurosci. Biobehav. Rev. 37, 1893–1916. https://doi.org/10.1016/j.neubiorev.2013.03.002.
Lavan, N., Short, B., Wilding, A., McGettigan, C., 2018. Impoverished encoding of speaker identity in spontaneous laughter. Evol. Hum. Behav. 39, 139–145. https://doi.org/10.1016/j.evolhumbehav.2017.11.002.
Leaver, A.M., Rauschecker, J.P., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J. Neurosci. 30, 7604–7612. https://doi.org/10.1523/JNEUROSCI.0296-10.2010.
Leaver, A.M., Rauschecker, J.P., 2016. Functional topography of human auditory cortex. J. Neurosci. 36, 1416–1428. https://doi.org/10.1523/JNEUROSCI.0226-15.2016.
Leaver, A.M., Van Lare, J., Zielinski, B., Halpern, A.R., Rauschecker, J.P., 2009. Brain activation during anticipation of sound sequences. J. Neurosci. 29, 2477–2485. https://doi.org/10.1523/JNEUROSCI.4921-08.2009.
Lemus, L., Hernández, A., Romo, R., 2009. Neural encoding of auditory discrimination in ventral premotor cortex. Proc. Natl. Acad. Sci. U. S. A. 106, 14640–14645. https://doi.org/10.1073/pnas.0907505106.
Leongómez, J.D., Mileva, V.R., Littleb, A.C., Roberts, S.C., 2017. Perceived differences in social status between speaker and listener affect the speaker's vocal characteristics. PLoS One 12. https://doi.org/10.1371/journal.pone.0179407.
Levréro, F., Carrete-Vega, G., Herbert, A., Lawabi, I., Courtiol, A., Willaume, E., Kappeler, P.M., Charpentier, M.J.E., 2015. Social shaping of voices does not impair phenotype matching of kinship in mandrills. Nat. Commun. 6 https://doi.org/10.1038/ncomms8609.
Lindblom, B., 1996. Role of articulation in speech perception: clues from production. J. Acoust. Soc. Am. 99, 1683–1692. https://doi.org/10.1121/1.414691.
Linville, S.E., 1996. The sound of senescence. J. Voice 10, 190–200. https://doi.org/10.1016/S0892-1997(96)80046-4.
Maguinness, C., Roswandowitz, C., von Kriegstein, K., 2018. Understanding the mechanisms of familiar voice-identity recognition in the human brain. Neuropsychologia 116, 179–193. https://doi.org/10.1016/j.neuropsychologia.2018.03.039.
Mantell, J.T., Pfordresher, P.Q., 2013. Vocal imitation of song and speech. Cognition 127, 177–202. https://doi.org/10.1016/j.cognition.2012.12.008.
Masaki, H., Tanaka, H., Takasawa, N., Yamazaki, K., 2001. Error-related brain potentials elicited by vocal errors. Neuroreport 12, 1851–1855. https://doi.org/10.1097/00001756-200107030-00018.
McAleer, P., Todorov, A., Belin, P., 2014. How do you say "hello"? Personality impressions from brief novel voices. PLoS One 9, e90779. https://doi.org/10.1371/journal.pone.0090779.
Mechelli, A., Price, C.J., Noppeney, U., Friston, K.J., 2003. A dynamic causal modeling study on category effects: bottom-up or top-down mediation? J. Cogn. Neurosci. 15, 925–934. https://doi.org/10.1162/089892903770007317.
Milesi, V., Cekic, S., Péron, J., Frühholz, S., Cristinzio, C., Seeck, M., Grandjean, D., 2014. Multimodal emotion perception after anterior temporal lobectomy (ATL). Front. Hum. Neurosci. 8, 275. https://doi.org/10.3389/fnhum.2014.00275.
Mileva, M., Tompkinson, J., Watt, D., Burton, A.M., 2018. Audiovisual integration in social evaluation. J. Exp. Psychol. Hum. Percept. Perform. 44, 128–138. https://doi.org/10.1037/xhp0000439.
Miller, C.T., Wang, X., 2006. Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. J. Comp. Physiol. A Neuroethol. Sensory Neural Behav. Physiol. 192, 27–38. https://doi.org/10.1007/s00359-005-0043-z.
Miller, C.T., Dimauro, A., Pistorio, A., Hendry, S., Wang, X., 2010. Vocalization induced cFos expression in marmoset cortex. Front. Integr. Neurosci. 1–15. https://doi.org/10.3389/fnint.2010.00128.
Miller, C.T., Thomas, A.W., Nummela, S.U., de la Mothe, L.A., 2015. Responses of primate frontal cortex neurons during natural vocal communication. J. Neurophysiol. 114, 1158–1171. https://doi.org/10.1152/jn.01003.2014.
Mitchell, R.L.C., 2007. fMRI delineation of working memory for emotional prosody in the brain: commonalities with the lexico-semantic emotion network. Neuroimage 36, 1015–1025. https://doi.org/10.1016/j.neuroimage.2007.03.016.
Möbes, J., Joppich, G., Stiebritz, F., Dengler, R., Schröder, C., 2008. Emotional speech in Parkinson's disease. Mov. Disord. 23, 824–829. https://doi.org/10.1002/mds.21940.
Monetta, L., Cheang, H.S., Pell, M.D., 2008. Understanding speaker attitudes from prosody by adults with Parkinson's disease. J. Neuropsychol. 2, 415–430. https://doi.org/10.1348/174866407X216675.
Muñoz, M., Mishkin, M., Saunders, R.C., 2009. Resection of the medial temporal lobe disconnects the rostral superior temporal gyrus from some of its projection targets in the frontal lobe and thalamus. Cereb. Cortex 19, 2114–2130. https://doi.org/10.1093/cercor/bhn236.
Munoz-Lopez, M.M., Mohedano-Moriano, A., Insausti, R., 2010. Anatomical pathways for auditory memory in primates. Front. Neuroanat. 4 https://doi.org/10.3389/fnana.2010.00129.
Murray, M.M., Camen, C., Gonzalez Andino, S.L., Bovet, P., Clarke, S., 2006. Rapid brain discrimination of sounds of objects. J. Neurosci. 26, 1293–1302. https://doi.org/10.1523/JNEUROSCI.4511-05.2006.
Nelson, L.R., Signorella, M.L., Botti, K.G., 2016. Accent, gender, and perceived competence. Hisp. J. Behav. Sci. 38, 166–185. https://doi.org/10.1177/0739986316632319.
Niedenthal, P.M., 2007. Embodying emotion. Science 316 (80-), 1002–1005. https://doi.org/10.1126/science.1136930.
O'Connor, J.J.M., Barclay, P., 2017. The influence of voice pitch on perceptions of trustworthiness across social contexts. Evol. Hum. Behav. 38, 506–512. https://doi.org/10.1016/j.evolhumbehav.2017.03.001.
O'Connor, J.J.M., Re, D.E., Feinberg, D.R., 2011. Voice pitch influences perceptions of sexual infidelity. Evol. Psychol. 9, 64–78. https://doi.org/10.1177/147470491100900109.
Oleszkiewicz, A., Pisanski, K., Lachowicz-Tabaczek, K., Sorokowska, A., 2017. Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. Psychon. Bull. Rev. 24, 856–862. https://doi.org/10.3758/s13423-016-1146-y.
Olson, I.R., Plotzker, A., Ezzyat, Y., 2007. The Enigmatic temporal pole: a review of findings on social and emotional processing. Brain 130, 1718–1731. https://doi.org/10.1093/brain/awm052.
Oveis, C., Spectre, A., Smith, P.K., Liu, M.Y., Keltner, D., 2016. Laughter conveys status. J. Exp. Soc. Psychol. 65, 109–115. https://doi.org/10.1016/j.jesp.2016.04.005.
Panksepp, J., 2003. Feeling the pain of social loss. Science 302 (80-), 237–239. https://doi.org/10.1126/science.1091062.
Pannese, A., Grandjean, D., Frühholz, S., 2016. Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions. Cortex 85, 116–125. https://doi.org/10.1016/j.cortex.2016.10.013.
Parkinson, A.L., Flagmeier, S.G., Manes, J.L., Larson, C.R., Rogers, B., Robin, D.A., 2012. Understanding the neural mechanisms involved in sensory control of voice production. Neuroimage 61, 314–322. https://doi.org/10.1016/j.neuroimage.2012.02.068.
Parkinson, A.L., Korzyukov, O., Larson, C.R., Litvak, V., Robin, D.A., 2013. Modulation of effective connectivity during vocalization with perturbed auditory feedback. Neuropsychologia 51, 1471–1480. https://doi.org/10.1016/j.neuropsychologia.2013.05.002.
Parsons, C.E., Young, K.S., Parsons, E., Stein, A., Kringelbach, M.L., 2012. Listening to infant distress vocalizations enhances effortful motor performance. Acta Paediatr. Int. J. Paediatr. 101, e189–91. https://doi.org/10.1111/j.1651-2227.2011.02554.x.
Parsons, C.E., Young, K.S., Joensson, M., Brattico, E., Hyam, J.A., Stein, A., Green, A.L., Aziz, T.Z., Kringelbach, M.L., 2014. Ready for action: a role for the human midbrain in responding to infant vocalizations. Soc. Cogn. Affect. Neurosci. 9, 977–984. https://doi.org/10.1093/scan/nst076.
Pasternak, T., Greenlee, M.W., 2005. Working memory in primate sensory systems. Nat. Rev. Neurosci. 6, 97–107. https://doi.org/10.1038/nrn1603.
Patel, S., Scherer, K.R., Björkner, E., Sundberg, J., 2011. Mapping emotions into acoustic space: the role of voice production. Biol. Psychol. 87, 93–98. https://doi.org/10.1016/j.biopsycho.2011.02.010.
Paulmann, S., Kotz, S.A., 2008. An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. Brain Lang. 105, 59–69. https://doi.org/10.1016/j.bandl.2007.11.005.
Paulmann, S., Ott, D.V.M., Kotz, S.A., 2011. Emotional speech perception unfolding in time: the role of the basal ganglia. PLoS One 6, e17694. https://doi.org/10.1371/journal.pone.0017694.
Pell, M.D., 2006. Chapter 17 Judging emotion and attitudes from prosody following brain damage. Prog. Brain Res. 156, 303–317. https://doi.org/10.1016/S0079-6123(06)56017-0.
Pell, M.D., 2007. Reduced sensitivity to prosodic attitudes in adults with focal right hemisphere brain damage. Brain Lang. 101, 64–79. https://doi.org/10.1016/j.bandl.2006.10.003.
Pell, M.D., Kotz, S.A., 2011. On the time course of vocal emotion recognition. PLoS One 6, e27256. https://doi.org/10.1371/journal.pone.0027256.
Pérez-Bellido, A., Anne Barnes, K., Crommett, L.E., Yau, J.M., 2018. Auditory frequency representations in human somatosensory cortex. Cereb. Cortex 28, 3908–3921. https://doi.org/10.1093/cercor/bhx255.
Pernet, C.R., Belin, P., 2012. The role of pitch and timbre in voice gender categorization. Front. Psychol. 3 https://doi.org/10.3389/fpsyg.2012.00023.
Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E.G., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. Neuroimage 119, 164–174. https://doi.org/10.1016/j.neuroimage.2015.06.050.
Péron, J., Dondaine, T., Le Jeune, F., Grandjean, D., Vérin, M., 2012. Emotional processing in parkinson's disease: a systematic review. Mov. Disord. 27, 186–199. https://doi.org/10.1002/mds.24025.

Péron, J., Frühholz, S., Ceravolo, L., Grandjean, D., 2015. Structural and functional connectivity of the subthalamic nucleus during vocal emotion decoding. Soc. Cogn. Affect. Neurosci. 11, 349–356. https://doi.org/10.1093/scan/nsv118.
Perrodin, C., Kayser, C., Logothetis, N.K., Petkov, C.I., 2011. Voice cells in the primate temporal lobe. Curr. Biol. 21, 1408–1415. https://doi.org/10.1016/j.cub.2011.07.028.
Perrodin, C., Kayser, C., Abel, T.J., Logothetis, N.K., Petkov, C.I., 2015. Who is that? Brain networks and mechanisms for identifying individuals. Trends Cogn. Sci. https://doi.org/10.1016/j.tics.2015.09.002.
Petkov, C.I., Jarvis, E.D., 2012. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. Front. Evol. Neurosci. 4 https://doi.org/10.3389/fnevo.2012.00012.
Petkov, C.I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., Logothetis, N.K., 2008. A voice region in the monkey brain. Nat. Neurosci. 11, 367–374. https://doi.org/10.1038/nn2043.
Petrides, M., Pandya, D.N., 1988. Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. J. Comp. Neurol. 273, 52–66. https://doi.org/10.1002/cne.902730106.
Pichon, S., Kell, C.A., 2013. Affective and sensorimotor components of emotional prosody generation. J. Neurosci. 33, 1640–1650. https://doi.org/10.1523/JNEUROSCI.3530-12.2013.
Pisanski, K., Mora, E.C., Pisanski, A., Reby, D., Sorokowski, P., Frackowiak, T., Feinberg, D.R., 2016. Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. Sci. Rep. 6 https://doi.org/10.1038/srep34389.
Poulson, C.L., Kymissis, E., Reeve, K.F., Andreatos, M., Reeve, L., 1991. Generalized vocal imitation in infants. J. Exp. Child Psychol. 51, 267–279. https://doi.org/10.1016/0022-0965(91)90036-R.
Prather, J.F., Peters, S., Nowicki, S., Mooney, R., 2008. Precise auditory-vocal mirroring in neurons for learned vocal communication. Nature 451, 305–310. https://doi.org/10.1038/nature06492.
Puts, D.A., Jones, B.C., Debruine, L.M., 2012. Sexual selection on human faces and voices. J. Sex Res. https://doi.org/10.1080/00224499.2012.658924.
Pye, A., Bestelmeyer, P.E.G., 2015. Evidence for a supra-modal representation of emotion from cross-modal adaptation. Cognition 134, 245–251. https://doi.org/10.1016/j.cognition.2014.11.001.
Rakić, T., Steffens, M.C., Mummendey, A., 2011a. When it matters how you pronounce it: the influence of regional accents on job interview outcome. Br. J. Psychol. 102, 868–883. https://doi.org/10.1111/j.2044-8295.2011.02051.x.
Rakić, T., Steffens, M.C., Mummendey, A., 2011b. Blinded by the accent! The minor role of looks in ethnic categorization. J. Pers. Soc. Psychol. 100, 16–29. https://doi.org/10.1037/a0021522.
Rathelot, J.A., Strick, P.L., 2009. Subdivisions of primary motor cortex based on cortico-motoneuronal cells. Proc. Natl. Acad. Sci. U. S. A. 106, 918–923. https://doi.org/10.1073/pnas.0808362106.
Rauschecker, J.P., 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear. Res. 271, 16–25. https://doi.org/10.1016/j.heares.2010.09.001.
Rauschecker, J.P., 2012. Ventral and dorsal streams in the evolution of speech and language. Front. Evol. Neurosci. 4, 7. https://doi.org/10.3389/fnevo.2012.00007.
Rauschecker, J.P., 2018. Where did language come from? Precursor mechanisms in nonhuman primates. Curr. Opin. Behav. Sci. 21, 195–204. https://doi.org/10.1016/j.cobeha.2018.06.003.
Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724. https://doi.org/10.1038/nn.2331.
Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 97, 11800–11806. https://doi.org/10.1073/pnas.97.22.11800.
Rigoulot, S., Pell, M.D., 2014. Emotion in the voice influences the way we scan emotional faces. Speech Commun. 65, 36–49. https://doi.org/10.1016/j.specom.2014.05.006.
Rigoulot, S., Fish, K., Pell, M.D., 2014. Neural correlates of inferring speaker sincerity from white lies: an event-related potential source localization study. Brain Res. 1565, 48–62. https://doi.org/10.1016/j.brainres.2014.04.022.
Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. Annu. Rev. Neurosci. 27, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230.
Rockwell, P., 2000. Lower, slower, louder: vocal cues of sarcasm. J. Psycholinguist. Res. 29, 483–495. https://doi.org/10.1023/A:1005120109296.
Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., Rauschecker, J.P., 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. 2, 1131–1136. https://doi.org/10.1038/16056.
Roy, S., Miller, C.T., Gottsch, D., Wang, X., 2011. Vocal control by the common marmoset in the presence of interfering noise. J. Exp. Biol. 214, 3619–3629. https://doi.org/10.1242/jeb.056101.
Roy, S., Zhao, L., Wang, X., 2016. Distinct neural activities in premotor cortex during natural vocal behaviors in a new world primate, the common marmoset (Callithrix jacchus). J. Neurosci. 36, 12169–12179. https://doi.org/10.1523/JNEUROSCI.1646-16.2016.
Sadagopan, S., Temiz-Karayol, N.Z., Voss, H.U., 2015. High-field functional magnetic resonance imaging of vocalization processing in marmosets. Sci. Rep. 5, 10950. https://doi.org/10.1038/srep10950.
Scherer, K.R., 1986. Vocal Affect Expression. A Review and a Model for Future Research. Psychol. Bull. 99, 143–165. https://doi.org/10.1037/0033-2909.99.2.143.
Scherer, K.R., Feldstein, S., Bond, R.N., Rosenthal, R., 1985. Vocal cues to deception: a comparative channel approach. J. Psycholinguist. Res. 14, 409–425. https://doi.org/10.1007/BF01067884.
Scheumann, M., Hasting, A.S., Kotz, S.A., Zimmermann, E., 2014. The voice of emotion across species: How do human listeners recognize animals' affective states? PLoS One 9, e91192. https://doi.org/10.1371/journal.pone.0091192.
Schirmer, A., Adolphs, R., 2017. Emotion perception from face, voice, and touch: comparisons and convergence. Trends Cogn. Sci. 21, 216–228. https://doi.org/10.1016/j.tics.2017.01.001.
Schirmer, A., Escoffier, N., 2010. Emotional MMN: anxiety and heart rate correlate with the ERP signature for auditory change detection. Clin. Neurophysiol. 121, 53–59. https://doi.org/10.1016/j.clinph.2009.09.029.
Schirmer, A., Zysset, S., Kotz, S.A., Von Cramon, D.Y., 2004. Gender differences in the activation of inferior frontal cortex during emotional speech perception. Neuroimage 21, 1114–1123. https://doi.org/10.1016/j.neuroimage.2003.10.048.
Schirmer, A., Chen, C.B., Ching, A., Tan, L., Hong, R.Y., 2013. Vocal emotions influence verbal memory: neural correlates and interindividual differences. Cogn. Affect. Behav. Neurosci. 13, 80–93. https://doi.org/10.3758/s13415-012-0132-8.
Schomers, M.R., Pulvermüller, F., 2016. Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. Front. Hum. Neurosci. 10 https://doi.org/10.3389/fnhum.2016.00435.
Schroeder, J., Epley, N., 2015. The sound of intellect: speech reveals a thoughtful mind, increasing a job candidate's appeal. Psychol. Sci. 26, 877–891. https://doi.org/10.1177/0956797615572906.
Schweinberger, S.R., Burton, A.M., 2003. Covert recognition and the neural system for face processing. Cortex 39, 9–30. https://doi.org/10.1016/S0010-9452(08)70071-6.
Schweinberger, S.R., Robertson, D.M.C., 2017. Audiovisual integration in familiar person recognition. Vis. cogn. 25, 589–610. https://doi.org/10.1080/13506285.2016.1276110.
Schweinberger, S.R., Herholz, A., Sommer, W., 1997. Recognizing famous voices: influence of stimulus duration and different types of retrieval cues. J. Speech Lang. Hear. Res. 40, 453–463. https://doi.org/10.1044/jslhr.4002.453.
Schweinberger, S.R., Casper, C., Hauthal, N., Kaufmann, J.M., Kawahara, H., Kloth, N., Robertson, D.M.C., Simpson, A.P., Zäske, R., 2008. Auditory adaptation in voice perception. Curr. Biol. 18, 684–688. https://doi.org/10.1016/j.cub.2008.04.015.
Schweinberger, S.R., Kloth, N., Robertson, D.M.C., 2011. Hearing facial identities: brain correlates of face-voice integration in person identification. Cortex 47, 1026–1037. https://doi.org/10.1016/j.cortex.2010.11.011.
Schweinberger, S.R., Kawahara, H., Simpson, A.P., Skuk, V.G., Zäske, R., 2014. Speaker perception. Wiley Interdiscip. Rev. Cogn. Sci. 5, 15–25. https://doi.org/10.1002/wcs.1261.
Scott, S.K., 2019. From speech and talkers to the social world: the neural processing of human spoken language. Science 366 (6461), 58–62. https://doi.org/10.1126/science.aax0288.
Scott, B.H., Mishkin, M., Yin, P., 2014. Neural correlates of auditory short-term memory in rostral superior temporal cortex. Curr. Biol. 24, 2767–2775. https://doi.org/10.1016/j.cub.2014.10.004.
Sehweinberger, S.R., 2001. Human brain potential correlates of voice priming and voice recognition. Neuropsychologia 39, 921–936. https://doi.org/10.1016/S0028-3932(01)00023-9.
Sei Jin, Ko, Judd, C.M., Stapel, D.A., 2009. Stereotyping based on voice in the presence of individuating information: vocal femininity affects perceived competence but not warmth. Personal. Soc. Psychol. Bull. 35, 198–211. https://doi.org/10.1177/0146167208326477.
Seyfarth, R.M., Cheney, D.L., Marler, P., 1980. Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. Science 210 (4471), 801–803. https://doi.org/10.1126/science.7433999.
Sidtis, J.J., Van Lancker Sidtis, D., 2003. A neurobehavioral approach to dysprosody. Semin. Speech Lang. 24, 93–105. https://doi.org/10.1055/s-2003-38901.
Silk, J.B., Seyfarth, R.M., Cheney, D.L., 2016. Strategic use of affiliative vocalizations by wild female baboons. PLoS One 11. https://doi.org/10.1371/journal.pone.0163978.
Simmonds, A.J., Leech, R., Iverson, P., Wise, R.J.S., 2014. The response of the anterior striatum during adult human vocal learning. J. Neurophysiol. 112, 792–801. https://doi.org/10.1152/jn.00901.2013.
Simonyan, K., Horwitz, B., 2011. Laryngeal motor cortex and control of speech in humans. Neuroscientist 17, 197–208. https://doi.org/10.1177/1073858410386727.
Skuk, V.G., Schweinberger, S.R., 2013. Adaptation aftereffects in vocal emotion perception elicited by expressive faces and voices. PLoS One 8. https://doi.org/10.1371/journal.pone.0081691.
Sliwa, J., Duhamel, J.R., Pascalis, O., Wirth, S., 2011. Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. Proc. Natl. Acad. Sci. U. S. A. 108, 1735–1740. https://doi.org/10.1073/pnas.1008169108.
Smiley, J.F., Falchier, A., 2009. Multisensory connections of monkey auditory cerebral cortex. Hear. Res. 258, 37–46. https://doi.org/10.1016/j.heares.2009.06.019.
Sodoyer, D., Rivet, B., Girin, L., Savariaux, C., Schwartz, J.-L., Jutten, C., 2009. A study of lip movements during spontaneous dialog and its application to voice activity detection. J. Acoust. Soc. Am. 125, 1184–1196. https://doi.org/10.1121/1.3050257.
Solis, M.M., Brainard, M.S., Hessler, N.A., Doupe, A.J., 2000. Song selectivity and sensorimotor signals in vocal learning and production. Proc. Natl. Acad. Sci. U. S. A. 97, 11836–11842. https://doi.org/10.1073/pnas.97.22.11836.
Stanley, D., 1931. The science of voice. Journal of the Franklin Institute. Oxford University Press, Oxford, UK, pp. 405–455. https://doi.org/10.1016/S0016-0032(31)90646-7.
Stephens, G.J., Silbert, L.J., Hasson, U., 2010. Speaker-listener neural coupling underlies successful communication. Proc. Natl. Acad. Sci. U. S. A. 107, 14425–14430. https://doi.org/10.1073/pnas.1008662107.
Suga, N., Ji, W., Ma, X., 2004. Criticisms of "Specific long-term memory traces in primary auditory cortex.". Nat. Rev. Neurosci. 5 https://doi.org/10.1038/nrn1366-c3, 1–1.

Sulpizio, S., Fasoli, F., Maass, A., Paladino, M.P., Vespignani, F., Eyssel, F., Bentler, D., 2015. The sound of voice: voice-based categorization of speakers' sexual orientation within and across languages. PLoS One 10, e0128882. https://doi.org/10.1371/journal.pone.0128882.

Tallal, P., 2004. Improving language and literacy is a matter of time. Nat. Rev. Neurosci. 5, 721–728. https://doi.org/10.1038/nrn1499.

Tigue, C.C., Borak, D.J., O'Connor, J.J.M., Schandl, C., Feinberg, D.R., 2012. Voice pitch influences voting behavior. Evol. Hum. Behav. 33, 210–216. https://doi.org/10.1016/j.evolhumbehav.2011.09.004.

Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., Kuriki, S., 2007. Neural correlates of auditory feedback control in human. Neuroscience 146, 499–503. https://doi.org/10.1016/j.neuroscience.2007.02.023.

Tremblay, S., Shiller, D.M., Ostry, D.J., 2003. Somatosensory basis of speech production. Nature 423, 866–869. https://doi.org/10.1038/nature01710.

Tuomainen, O., Hazan, V., 2016. Suprasegmental characteristics of spontaneous speech produced in good and challenging communicative conditions by younger and older adults. J. Acoust. Soc. Am. 140 https://doi.org/10.1121/1.4971112, 3444–3444.

Van Lancker, D., Kreiman, J., Emmorey, K., 1985. Familiar voice recognition: patterns and parameters Part I: recognition of backward voices. J. Phon. 13, 19–38. https://doi.org/10.1016/s0095-4470(19)30723-5.

Van Lancker Sidtis, D., Pachana, N., Cummings, J.L., Sidtis, J.J., 2006. Dysprosodic speech following basal ganglia insult: toward a conceptual framework for the study of the cerebral representation of prosody. Brain Lang. 97, 135–153. https://doi.org/10.1016/j.bandl.2005.09.001.

von Holst, E., Mittelstaedt, H., 1950. Das Reafferenzprinzip - wechselwirkungen zwischen zentralnervensystem und peripherie. Naturwissenschaften 37, 464–476. https://doi.org/10.1007/BF00622503.

Von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A.L., Kell, C.A., Grüter, T., Kleinschmidt, A., Kiebel, S.J., 2008. Simulation of talking faces in the human brain improves auditory speech recognition. Proc. Natl. Acad. Sci. U. S. A. 105, 6747–6752. https://doi.org/10.1073/pnas.0710826105.

Von Kriegstein, K., Smith, D.R.R., Patterson, R.D., Kiebel, S.J., Griffiths, T.D., 2010. How the human brain recognizes speech in the context of changing speakers. J. Neurosci. 30, 629–638. https://doi.org/10.1523/JNEUROSCI.2742-09.2010.

Walsh, B., Smith, A., 2012. Basic parameters of articulatory movements and acoustics in individuals with Parkinson's disease. Mov. Disord. 27, 843–850. https://doi.org/10.1002/mds.24888.

Wambacq, I.J.A., Shea-Miller, K.J., Abubakr, A., 2004. Non-voluntary and voluntary processing of emotional prosody: an event-related potentials study. Neuroreport 15, 555–559. https://doi.org/10.1097/00001756-200403010-00034.

Warren, J.E., Sauter, D.A., Eisner, F., Wiland, J., Dresner, M.A., Wise, R.J.S., Rosen, S., Scott, S.K., 2006. Positive emotions preferentially engage an auditory-motor "mirror" system. J. Neurosci. 26, 13067–13075. https://doi.org/10.1523/JNEUROSCI.3907-06.2006.

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., Belin, P., 2014. Crossmodal adaptation in right posterior superior temporal sulcus during face-voice emotional integration. J. Neurosci. 34, 6813–6821. https://doi.org/10.1523/JNEUROSCI.4478-13.2014.

Wattendorf, E., Westermann, B., Fiedler, K., Kaza, E., Lotze, M., Celio, M.R., 2013. Exploration of the neural correlates of ticklish laughter by functional magnetic resonance imaging. Cereb. Cortex 23, 1280–1289. https://doi.org/10.1093/cercor/bhs094.

Weston, P.S.J., Hunter, M.D., Sokhi, D.S., Wilkinson, I.D., Woodruff, P.W.R., 2015. Discrimination of voice gender in the human auditory cortex. Neuroimage 105, 208–214. https://doi.org/10.1016/j.neuroimage.2014.10.056.

Wild, C.J., Linke, A.C., Zubiaurre-Elorza, L., Herzmann, C., Duffy, H., Han, V.K., Lee, D.S. C., Cusack, R., 2017. Adult-like processing of naturalistic sounds in auditory cortex by 3- and 9-month old infants. Neuroimage 157, 623–634. https://doi.org/10.1016/j.neuroimage.2017.06.038.

Wilkins, M.R., Seddon, N., Safran, R.J., 2013. Evolutionary divergence in acoustic signals: causes and consequences. Trends Ecol. Evol. 28, 156–166. https://doi.org/10.1016/j.tree.2012.10.002.

Williams, H., Nottebohm, F., 1985. Auditory responses in avian vocal motor neurons: a motor theory for song perception in birds. Science 229 (4710), 279–282. https://doi.org/10.1126/science.4012321.

Yeterian, E.H., Pandya, D.N., 1998. Corticostriatal connections of the superior temporal region in rhesus monkeys. J. Comp. Neurol. 399, 384–402. https://doi.org/10.1002/(SICI)1096-9861(19980928)399:3<384::AID-CNE7>3.0.CO;2-X.

Young, A., 2016. Facial expression recognition: selected works of Andy Young. Facial Expression Recognition: Selected Works of Andy Young, pp. 1–329. https://doi.org/10.4324/9781315715933.

Yovel, G., O'Toole, A.J., 2016. Recognizing people in motion. Trends Cogn. Sci. 20, 383–395. https://doi.org/10.1016/j.tics.2016.02.005.

Zäske, R., Schweinberger, S.R., 2011. You are only as old as you sound: auditory aftereffects in vocal age perception. Hear. Res. 282, 283–288. https://doi.org/10.1016/j.heares.2011.06.008.

Zäske, R., Schweinberger, S.R., Kaufmann, J.M., Kawahara, H., 2009. In the ear of the beholder: neural correlates of adaptation to voice gender. Eur. J. Neurosci. 30, 527–534. https://doi.org/10.1111/j.1460-9568.2009.06839.x.

Zäske, R., Schweinberger, S.R., Kawahara, H., 2010. Voice aftereffects of adaptation to speaker identity. Hear. Res. 268, 38–45. https://doi.org/10.1016/j.heares.2010.04.011.

Zäske, R., Volberg, G., Kovács, G., Schweinberger, S.R., 2014. Electrophysiological correlates of voice learning and recognition. J. Neurosci. 34, 10821–10831. https://doi.org/10.1523/JNEUROSCI.0581-14.2014.

Zhang, C., Tan, T., 2008. Voice disguise and automatic speaker recognition. Forensic Sci. Int. 175, 118–122. https://doi.org/10.1016/j.forsciint.2007.05.019.

Zhao, L., Rad, B.B., Wang, X., 2019. Long-lasting vocal plasticity in adult marmoset monkeys. Proc. R. Soc. B Biol. Sci. 286 https://doi.org/10.1098/rspb.2019.0817.

Ziegler, W., Ackermann, H., 2017. Subcortical contributions to motor speech: phylogenetic, developmental. Clinical. Trends Neurosci. 40, 458–468. https://doi.org/10.1016/j.tins.2017.06.005.

Zürcher, Y., Willems, E.P., Burkart, J.M., 2019. Are dialects socially learned in marmoset monkeys? Evidence from translocation experiments. PLoS One 14. https://doi.org/10.1371/journal.pone.0222486.